



Understanding Digital Well-being in Social Media Use: App-Specific Use, Goal Pursuit, and Platform Designs

Citation

Liu, Yuning. 2026. Understanding Digital Well-being in Social Media Use: App-Specific Use, Goal Pursuit, and Platform Designs. Doctoral Dissertation, Harvard University Graduate School of Arts and Sciences.

Link

<https://dash.harvard.edu/handle/1/42740408>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.

Please share how this access benefits you. [Submit a story](#)

HARVARD

Kenneth C. Griffin



GRADUATE SCHOOL
OF ARTS AND SCIENCES

THESIS ACCEPTANCE CERTIFICATE

The undersigned, appointed by the
Department of Population Health Sciences,
have examined a dissertation, entitled,

**“Understanding Digital Well-being in Social Media Use:
App-Specific Use, Goal Pursuit, and Platform Designs”**

presented by

YUNING LIU

candidate for the degree of Doctor of Philosophy
and hereby certify that it is worthy of acceptance.

David Canning

David Canning (May 11, 2026 11:31:43 EDT)

Dr. David Canning, Ph.D., Committee Chair, Harvard T.H. Chan School of Public Health

Amit

Amit Goldenberg (May 11, 2026 16:35:24 EDT)

Dr. Amit Goldenberg, Ph.D., Harvard Business School

K. Viswanath

K. Viswanath (May 11, 2026 16:45:50 EDT)

Dr. Kasisomayajula Viswanath, Ph.D., Harvard T.H. Chan School of Public Health

Date: 05 May 2026

Understanding Digital Well-being in Social Media Use: App-Specific Use, Goal Pursuit, and Platform Designs

Yuning Liu

A dissertation submitted to the faculty of
The Harvard T. H. Chan School of Public Health
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in Population Health Sciences
in the Department of Global Health and Population

Harvard University
Cambridge, Massachusetts

May 2026

© 2026 Yuning Liu
All rights reserved.

Understanding Digital Well-being in Social Media Use: App-Specific Use, Goal Pursuit, and Platform Designs

Abstract

Social media platforms are now deeply embedded in everyday life, shaping how people communicate, seek information, and experience social connection. Over the past two decades, a large and rapidly growing body of research has examined the relationship between social media use and users' health and well-being. Findings from the digital well-being literature have remained mixed, documenting both positive and negative associations between social media use and well-being. These mixed findings are partly attributable to the complexity and heterogeneity of social media use, which involves highly content- and context-dependent behaviors, and are further compounded by persistent limitations in access to behavioral data from online platforms.

This dissertation contributes to the literature on social media use and well-being through three correlated and complementary perspectives. **The first chapter** reviews existing research on social media and well-being, outlines key conceptual and methodological challenges in the field, and introduces the three following empirical chapters, explaining how they are positioned within the existing evidence and how they contribute to the field. **The second chapter** introduces a novel approach to measuring objective social media use using user-donated screenshots and examines how app-specific usage patterns relate to daily well-being outcomes. **The third chapter** shifts from exposure-based measures to psychological processes of goal pursuit, examining how goal awareness and goal strength shape momentary social media experiences and well-being, and incorporating TikTok data donation packages to advance objective measurement. **The fourth chapter** shifts attention to the platform level, using Large Language Model (LLM)-assisted content analysis to examine how major social media companies articulate and signal compliance

with digital well-being practices, providing empirical evidence relevant to policy and regulatory discussions. Together, these studies offer a multi-dimensional account of how social media use, user psychology, and platform design jointly shape well-being in digital environments.

The second chapter investigates the association between app-specific social media use and daily well-being among adolescents using app specific use data extracted from screenshots. Drawing on a two-week ecological momentary assessment (EMA) on social media use and well-being, combined with user-donated screenshots from 374 Swiss adolescents (mean age = 15.71; 61% female), the study introduces “Appnomes,” a set of app-level metrics capturing screen time, number of activations, and notifications across platforms. These behavioral indicators were linked to daily measures of hedonic and eudaimonic well-being using multilevel models. Overall, results revealed small, mixed, or null associations between app-specific use and well-being. While certain associations emerged, such as longer TikTok use predicting lower same-day eudaimonic well-being but higher next-day positive emotions, Snapchat screen time predicting lower relaxation, competence, and positive affect, and WhatsApp activations predicting greater feelings of connection. Collectively, these results challenge the moral panic on the detrimental impact of social media use on teens well-being and underscore the limitations of exposure-based models that ignore contextual and psychological factors.

The third chapter shifts focus from behavioral exposure to underlying psychological processes, examining how goal awareness, defined as awareness of one’s goal for social media use, and goal strength, the degree to which an individual is motivated to pursue that goal, shape momentary user experiences and well-being. Using a 7-day EMA design with 266 U.S. adults (six surveys per day; 71.3% response rate), the study integrates self-reported goals for social media use and well-being outcomes, screentime screenshots, and user-donated TikTok behavioral data. Within individuals, higher goal awareness and goal strength were associated with more positive experiences and better well-being, respectively, particularly when social media use served goals such as coping, entertainment, passing time, or habitual use. TikTok data showed little behavioral

variation by goal awareness and goal strength, except that users with higher goal awareness skipped videos faster. Findings highlight goal awareness as a key process shaping momentary social media use experiences, while also underscore the need to better capture behavioral signals of intentional social media use.

The fourth chapter extends the analysis from users to platforms, examining social media platforms' claims to promote digital well-being practices. Digital media platforms increasingly face pressure from the public and legislators worldwide to mitigate harm to user well-being and promote healthier digital environments. Platforms appear to be paying greater attention to digital well-being proclaiming actions being taken by them to mitigate the harm. This study provides arguably the first large-scale analysis of digital well-being related-practices claimed by platforms using LLM-assisted content analysis of 13,497 company announcements published by seven major platforms (Facebook, Instagram, WhatsApp, X/Twitter, TikTok, YouTube, and Snapchat) from each platform's first post through May 15, 2025. Our analysis identified 5,685 distinct digital well-being practices, showing a steady increase in platform-reported efforts since 2019, with a notable surge following the introduction of the General Data Protection Regulation in the European Union. Across the four identified mechanisms of digital well-being practices, more than 80 percent of practices fall within content-based governance and user-initiated controls, with far fewer addressing default safety features or personalized algorithmic curation. Only a small share of announcements includes user-level impact evaluation, leaving the implementation status and actual effects of these practices largely unknown. The study indicates that platforms appear responsive to public and legislative pressures, but the actions they highlight tend to emphasize visible, content-focused measures rather than deeper changes to underlying algorithmic operations.

Taken together, this dissertation highlights the complexity of the relationship between social media use and well-being. It addresses this complexity through two complementary structural approaches: examining the psychological processes underlying social media use to

conceptualize use as goal pursuit processes and extending the scope of digital well-being to the platform level by analyzing platform design practices. The insights from this dissertation are particularly relevant for emerging AI-mediated tools, where both user psychology and platform design choices play central roles in shaping user experiences and well-being. Finally, this dissertation advances methodological approaches by integrating objective social media data, collected through screenshots and platform data donation packages, with self-reported measures to enable more rigorous study of digital well-being.

Table of Contents

Title page	i
Copyright.....	ii
Abstract.....	iii
Table of contents	vii
List of Figures	ix
List of Tables.....	x
Glossary of Terms.....	xi
Dedication	xii
Acknowledgement.....	xiii
Chapter 1: Introduction	1
1.1 Overview	1
1.2 Literature review: social media use and well-being.....	2
1.3 Challenges in digital well-being studies.....	7
1.4 Overview of the empirical chapters	9
Chapter 2: Appname analysis reveals small or no associations between social media app-specific usage and adolescent well-being.....	11
2.1 Introduction.....	12
2.2 Method	21
2.3 Results	30
2.4 Discussion	42
Supplementary materials.....	49
Figures.....	49
Tables	57
Chapter 3: Higher goal awareness improves social media user experience and well-being	60
3.1 Introduction.....	61
3.2 Method	67
3.3 Results	84
3.4 Discussion	100
Supplementary materials.....	112
Figures.....	112
Tables	129
Codebook	138

Chapter 4 – Digital well-being features on social media platforms: an LLM-assisted analysis of social media company announcement.....	144
4.1 Introduction.....	145
4.2 Method	150
4.3 Results	155
4.4 Discussion	178
Supplementary materials.....	187
Figures.....	187
Tables	195
Codebook	198
Prompts	205
Chapter 5: Conclusion	220
5.1 Key Findings and Implications.....	220
5.2 Future Research.....	222
Reference	225

List of Figures

Figure 1.1 An example of Appnomo on the screentime (A), number of activations (B), and number of notifications (C), of one female (1) and one male (2) study participant in the HappyB study	25
Figure 1.2. The distribution of person-level average screentime (A), number of activations (B), and number of notifications (C) by gender in the HappyB study.....	32
Figure 1.3. Correlation between average screentime (A), number of activations (B), and number of notifications (C) of selected Apps and average hedonic and eudaimonic well-being in the between-person level in the HappyB study.....	33
Figure 1.4. Forest plot of the mixed linear models results on the associations between TikTok, WhatsApp, and Instagram screentime, number of activations, and number of notifications and hedonic/eudaimonic well-being in the HappyB study.....	36
Figure 1.5. Associations of TikTok, WhatsApp, and Instagram usage with hedonic and eudaimonic well-being outcomes on the same day, the following day (1-day lag), and two days later (2-day lag) in the HappyB study.....	38
Figure 2.1. The study design for the goal awareness and social media experience study	67
Figure 2.2. The example time of TikTok activities in the episodes matched with EMA survey for two participants	74
Figure 2.3. Associations of goal awareness and goal strength with happiness, by subgroup of momentary goal types.....	93
Figure 2.4. Within-person (A) and between-person (B) associations between TikTok use behaviors and goal awareness, goal strength, social media use experience, and subjective well-being	95
Figure 2.5. Distribution of changes in social media use agency (A) and valence (B) mindsets..	99
Figure 3.1. Introductions of digital wellbeing practices for platforms from 2008-2025	156
Figure 3.2. Target user groups of digital well-being practices by platforms	158
Figure 3.3 Distribution of digital well-being practices in TikTok	159
Figure 3.4. Author positions in the company announcements with digital well-being practices	160
Figure 3.5. Targeted harms of the digital well-being practices in company announcement	161
Figure 3.6. Topical focus of the digital wellbeing practices from company announcements.....	162
Figure 3.7. Mechanisms of the digital wellbeing practices from company announcements.....	165
Figure 3.8. Evaluation status of the digital wellbeing practices from company announcements	175
Figure 3.9. Evaluation status of the digital wellbeing practices from company announcements	177

List of Tables

Table 1.1. Mixed linear model results on the standardized associations between screentime, number of activations, and number of notifications and hedonic/eudaimonic well-being in the HappyB study.....	34
Table 2.1. Between-person and within-person associations between social media motivation awareness, strength, use experience, and well-being	91
Table 3.1. Selected examples of policies addressing digital well-being and online safety	147

Glossary of Terms

Disability-Adjusted Life Years (DALYs)

General Data Protection Regulation (GDPR)

Digital Services Act (DSA)

European Union (EU)

World Health Organization (WHO)

Ecological Momentary Assessment (EMA)

Standard Deviation (SD)

Application Programming Interface (API)

Optical Character Recognition (OCR)

Open Science Framework (OSF)

Generalized Anxiety Disorder–2 (GAD-2)

Patient Health Questionnaire–2 (PHQ-2)

Large Language Model (LLM)

Akaike Information Criterion (AIC)

Root Mean Squared Error (RMSE)

Age-Appropriate Design Code Act (AADC)

Intraclass Correlation Coefficient (ICC)

Food and Drug Administration (FDA)

Dedication

To my mother, Yanzhen Qiao, whose belief in me has carried me through;
to my father, Zheng Liu, who opened my eyes to the world;
to my grandparents, Cuilin Yang and Fuliang Qiao, who raised me with unconditional love;
to my husband, Sen Yang, for his trust, support, commitment, and love;
and to my younger and older selves: please keep moving forward and exploring life with
courage and curiosity

Acknowledgement

This dissertation would not exist without the extraordinary community of mentors, collaborators, friends, and family who have walked this journey with me.

My committee members

I want to thank my dissertation advisor, Professor David Canning, who has seen me through every stage of this work. I still remember sitting in his office during my first year, insisting that I wanted to study social media because it mattered, even though I did not yet know how exactly to study it. David has witnessed my growth from that moment on, and his steady guidance, patience, and trust gave me the freedom to follow my passion.

I am profoundly grateful to Professor Kasisomayajula ("Vish") Viswanath, whose interdisciplinary insights across fields opened entirely new doors for me. Every conversation with Vish brought a new insight or a new question, and his expertise in how to make real-world, population-level change that reduces social disparities, has shaped how I think about this field.

I am also deeply grateful to Professor Amit Goldenberg, who decided to work with me even when my ideas were still rough and unformed. From Amit I learned how to become a scientific researcher: how to ask deep questions, how to do empirical work rigorously, how to collaborate, how to learn from others, and how to guide others in turn. I could not have continued this work without him. Amit empowered me to think of myself as a researcher, and that has made all the difference.

My collaborators on the dissertation chapters

I want to thank my collaborators for the three chapters of this dissertation. Professor Laura Marciano was the first researcher to generously share her resources with me, and she ended an anxious early period of my PhD when I did not yet know whom to work with or how to begin real

empirical work. Laura introduced me to the fields of mobile communication and media psychology and gave me the freedom to find my own footing within them.

For the second chapter, I am grateful to Professor Yael Millgram, who shared her resources and guided me through my first-ever experience of ecological momentary assessment data collection, and to Professor Joseph Bayer, who agreed to advise this work, guided me through theories of mobile media habit, and welcomed me into his lab, which is becoming a second intellectual home during my PhD journey.

For the third chapter, I thank Dr. Ravi Iyer, who thought of me for so many different projects, cared deeply about what I wanted to achieve, and empowered me to explore the field of social media well-being and human-centered AI, not only through empirical analysis but through cross-sector collaborations that connected me with industry. I am also grateful to the two undergraduate students who have since become close friends, Stanley Huang and Emily Xing, who were patient through my own exploration of analytic methods and of what it means to be a mentor. They worked diligently and produced far more learning and results than I expected. I feel I owe them a great deal, and I hope they get the best of everything in what comes next.

My mentors and broader research community

Beyond the dissertation itself, I had the great fortune of working with Amanda Yarnell and Professor Matthew Motta on projects supporting content creators' science communication on social media platforms. This work did not make it into the dissertation, but it deepened my understanding of the different stakeholders in the digital well-being ecosystem.

I also want to thank Professor Bryn Austin, along with Jill Kavanaugh and Professor Zachary Ward, who brought me into the world of simulation modeling for platform design and health outcomes. It is through that collaboration that I came to understand the dark patterns and design features on social media platforms that structure user experience.

I owe a particular thank-you to Professor Jeffrey Hall, with whom I worked as a research assistant during his year at the Berkman Klein Center. Jeff taught me, from beginning to end, what the ecology of social media experience really means, and he believed in me and trusted me even when I was only beginning to explore this field.

I am also grateful to Professor Lucía Magis-Weinberg, Professor Nathaniel Osgood, Professor Nathanael Fast, and my mentor and dear friend Professor Naijia Liu, whose mentorship and generosity have meant a great deal to me.

My peers

To my peers and friends in the Goldenberg Lab — Yaoxi Shi, Hanqiu Li, Federico Zimmerman, Yajun Cao, Sofie Labat, Cathy Fang, Noa Cohen Eick, Matan Eshed, Geshi Yeung, Sumer Vaid, and Catherine Huanh — thank you for the everyday intellectual support and social companionship.

To my PhD cohort friends Sun Kim, Yuchen Liu, and Sian Tsuei, thank you for sharing your experience, concerns, and insights, and walking through this PhD together.

To the researchers and the content-creator community at the Center for Health Communication, thank you for grounding me in another intellectual home and for reminding me that research is also a practice, lived out in the real world.

I have also been lucky to work with an extraordinary group of undergraduate research assistants: Emily Xing, Stanley Huang, Darwin Li, Jeffrey Zhou, Natalia Siwek, Ziqi Zheng, and Zein Alabdeen Zammar. Your curiosity, energy, and care made the work better.

Funding

I gratefully acknowledge funding from the Harvard Data Science Initiative, the Harvard Lee Kum Sheung Center for Health and Happiness, the Institute for Quantitative Social Science, the Harvard Center for Population and Development Studies, and the Department of Global

Health and Population at the Harvard T.H. Chan School of Public Health. All of these supports made my dissertation research possible.

A special thanks to Barbara Heil

I genuinely do not know how I would have survived the billion small steps of this PhD without you. Thank you for your kindness, support, and patience with me at every turn.

Family and friends

Finally, to my family — my parents, my grandparents, my aunt, and my husband — thank you for your love and support, and for being my home. And to the friends I met outside of academia — my roommates Jiayin Qu, Jinyu Li, and Hui Ren, my dear friend Yiyun (Star) Chen, and my friends from the climbing gyms — thank you for keeping me human through all of this. This is for all of you.

Chapter 1: Introduction

1.1 Overview

Over the past two decades, social media platforms have transformed not only the landscape of human communication but also everyday life. Unlike prior generations of mass media, which operated primarily through one-to-many broadcast models, social media enables many-to-many interaction, positioning individuals as consumers, creators, distributors, and curators of content within digital networks [1,2]. Consequently, social media has become increasingly woven into daily routines. At the same time, algorithmic recommendation systems personalize information flows at scale, tailoring content to maximize individual attention and engagement [3,4]. This continuous personalization hooks some users into prolonged engagement [5], resulting in widely reported experiences of information and connection overload due to fragmented exposure [6], depletion of self-control [7], and increased exposure to harmful content [8], with the extent of these experiences varies considerably across individuals and contexts.

Parallel to the rise of social media platforms, there has been a widely observed crisis in mental health and digital well-being. Globally, over one billion people were living with mental health conditions in 2025 [9], with a continuing upward trend [10]. The trend is particularly striking among children and adolescents. One in seven individuals aged 10–19 experiences a mental disorder, accounting for 15% of the global burden of disease in this age group [11]. The number of disability-adjusted life years (DALYs) attributable to mental disorders among those aged 10–24 increased from 26.1 million in 1990 to 36.3 million in 2021 globally [12]. In the United States, nearly one in five children aged 3–17 (21%) has been diagnosed with a mental, emotional, or behavioral condition [13].

The parallel increases in social media use and mental health concerns have heightened attention to digital well-being, defined as individuals' overall experiences and well-being outcomes in relation to their use of digital technologies. Since 2023, more than 10% of adults report negative

experiences or seeing content that they think are bad for the world on platforms such as X (formerly Twitter), Facebook, Instagram, YouTube, and TikTok [14]. For parents who themselves face challenges in digital well-being, additional responsibilities arise in supporting children's digital experiences, creating further challenges for families and schools working with youth [15,16]. As emphasized in the U.S. Surgeon General's Advisory, more research is needed to clarify the mechanisms behind platform use and health outcomes [17].

In response to the crisis in digital well-being, a substantial body of research has examined the impact of social media on well-being and health. Over the past two decades, the central research question has shifted from "Is social media use good or bad?" to "Under what circumstances, for whom, and what types of use are harmful or beneficial?" [18–20] More recently, discussions have expanded to governance questions concerning the evidence needed to inform platform design [21], industry regulation [22], and global legislation [23,24]. Although substantial progress has been made in improving digital well-being, concerns of users' mental health remain prevalent (WHO teens 2024) and regulatory frameworks are still evolving [11,25], indicating the need for continued empirical research to guide effective intervention and governance. This dissertation contributes to this broader effort by providing evidence on how social media use, user psychology, and platform design relate to user well-being. In this chapter, I first review the literature on social media and well-being, then summarize key barriers in studying digital well-being, and introduce the three empirical chapters and their contributions.

1.2 Literature review: social media use and well-being

A rapidly expanding body of research has investigated whether, and under what conditions, social media use is related to well-being and mental health. This literature spans multiple disciplines, including psychology, communication, public health, computer science, and human computer interaction, and reflect diverse theoretical and methodological approaches.

Time of use

Despite the common lay perception that social media use is inherently harmful, empirical research examining the relationship between social media use and well-being has yielded mixed findings. Systematic reviews indicate that associations between social media use and well-being are generally small and heterogeneous due to differences in study populations, designs, disciplinary approaches, platforms and well-being outcomes examined [26,27]. Some meta-analyses document statistically significant negative associations between overall social media use and well-being, but with small effect sizes ranging approximately from -0.05 to -0.15 standard deviation [18,28]. For example, one review estimated that overall social media use explains roughly 1% of the variance in well-being [29]. At the same time, other meta-analysis also suggests that certain dimensions of well-being, such as hedonic well-being (r , Pearson correlation coefficient = 0.11) and social well-being ($r = 0.07$), are positively associated with social media use [30]. Notably, inconsistency in this literature extends beyond the estimated associations between social media use and well-being to disagreements in their interpretation, with similar small-to-moderate effects alternatively described as modest or as serious by different researchers [20]. Together, these mixed findings underscore the heterogeneity of the relationship between social media use and well-being and the complexity of social media experiences. Moreover, the mixed findings point to the need to move beyond questions of overall use toward a more nuanced understanding of how social media is used and experienced.

Passive-active social media use

Beyond overall time spent on social media, a major line of research has examined what users actually do on social media platforms and how these activities relate to well-being. One widely used framework distinguishes between active social media use, which involves interacting with others (e.g., messaging or posting content) [18], and passive use, which involves consuming social media content without direct interaction [31]. Early hypotheses and empirical studies linked

active social media use to more positive well-being outcomes and passive use to more negative ones [32,33]. However, recent reviews indicate little support for a net negative association between passive use and well-being, and suggest that passive use can also involve positive experiences such as enjoyment, inspiration, and online support [18,34,35]. Moreover, the active–passive dichotomy has been criticized as overly simplistic and conceptually imprecise, as passive use collapses diverse experiences and individual heterogeneity [20,34].

Activities on social media

Moving beyond the active–passive framework, research on specific social media activities, such as posting or browsing, has also yielded mixed findings regarding well-being. For example, posting behavior has been associated with poorer academic performance in some studies [36], but with improved affective well-being in others [37], particularly when posting involves authentic self-expression [38]. Similarly, browsing activities have been associated with decreased well-being in some contexts [37], especially when browsing others’ or influencers’ content, but not when browsing one’s own profile [39], with negative social comparison as a key mediating mechanism [15]. With the rise of short-form video platforms, feed browsing has become an increasingly prominent activity, with studies documenting negative effects of doomscrolling [40,41], alongside evidence that browsing certain types of content, such as videos depicting prosocial behavior, can have positive effects [42], and that browsing relaxing videos may serve as a coping strategy for everyday stress [43].

Content on social media

Beyond the activities people engage in, a growing body of research has focused on the content users are exposed to on social media and its implications for well-being. A substantial body of research has examined exposure to harmful social media content, including misinformation [44], hate speech [45], and cyberbullying [46], and linked such exposure to poorer

well-being outcomes, such as increased psychological distress and anxiety. Some forms of harmful content disproportionately affect specific subpopulations, such as thinspiration content among adolescent girls and misogynistic content within online incel communities [47,48]. At the same time, other studies highlight the benefits of positive social media content for well-being, including community building among individuals with mental health conditions [49], peer support within self-harm communities [50], and exposure to creative or inspiring content [51]. Together, this content-focused literature underscores the heterogeneous nature of social media experiences, as the same amount of screen time or the same activity can expose users to content with vastly different psychological implications.

Other dimensions of social media use

Beyond activities and content exposure, researchers have further dissected social media use along several additional dimensions, including use heterogeneity across platforms, users' subjective experiences, contextual factors, and behavioral patterns, each adding layers of complexity to understanding its relationship with well-being. One line of work focuses on users' subjective experiences, including measures such as the Digital Stress Scale, which captures experiences like fear of missing out, information overload, and online vigilance [52], and the Social Media Mindset Scale, which assesses users' perceived agency and valence in their own social media use [53]; both have been shown to predict well-being. Researchers have also emphasized contextual factors, such as the physical and social settings in which social media use occurs [54], and their interaction with behavioral patterns, including variability, fragmentation, and burstiness of use [55,56]. Taken together, these approaches underscore that social media use is deeply woven into daily life and is inherently high-dimensional, context-dependent, and content-dependent.

Temporality of social media use and its affective responses is another high-dimensionality problem in digital well-being research. Social media exposure is not a static quantity but a process

that unfolds across time. The temporal structure of social media exposure, such as when use occurs, how it clusters across the day, whether effects accumulate gradually or spike and rapidly recede, is a substantive dimension of exposure that existing measures rarely capture fully. Critically, the effects of social media use on well-being may operate at fundamentally different timescales: some responses are short-lived and affective, tied to a specific interaction, while others are long-term and cumulative, reflecting the slow accretion of impacts over months or years. The temporal complexity of social media exposure means that studies often cannot make explicit at which temporal level their measures and inferences are operating. The mismatch between the granularity of exposure measurement and outcome assessment compounds this ambiguity. Objective behavioral data may capture app opens at the resolution of seconds or minutes, or aggregate use into daily totals, while well-being is typically assessed at intervals of hours or days, as shown in Chapter 2. Short-lived affective responses to specific interactions may have already receded by the time a well-being measure is taken. At the other end of the spectrum, long-term cumulative effects remain understudied, in large part because tracking participants across months or years is logistically and financially costly.

Even within the short-term window, the choice of lag structure, as examined in Chapters 2 and 3, introduces assumptions about when effects manifest that may not hold uniformly across individuals, contexts, or platforms. The present dissertation focuses on short-term within-person associations, but does not resolve the full complexity of temporal dynamics in social media exposure. This remains an open methodological frontier requiring progress on two fronts: better measurement tools to align the temporal resolution of exposure and outcome data, and stronger theoretical frameworks to guide assumptions about the timescales over how social media use and well-being co-evolve.

In summary, the high degree of context and content dependence of social media use poses great challenges in studying its impact on well-being. Unlike comparably more discrete behaviors such as smoking or exercising, social media use is deeply woven into nearly all aspects

of daily life and cannot be adequately characterized by a small set of behavioral indicators. For any individual at any moment, social media use reflects a complex combination of personal characteristics, platform features, situational context, activities and content exposures, and emotional or cognitive responses.

1.3 Challenges in digital well-being studies

The mixed findings in the relationship between social media use and well-being also reflects fundamental methodological and structural challenges in studying social media as a behavioral exposure. First, causal identification in the study of social media and well-being is particularly challenging because social media exposure is high-dimensional, rapidly evolving, and deeply embedded in everyday life. Even when holding time spent constant, individuals may engage in different activities, encounter different content, and allocate attention in distinct ways, often within seconds [27,57,58]. Many of these dimensions, such as the content viewed, are difficult to observe or measure comprehensively. As a result, empirical models may omit important aspects of the underlying exposure, introducing endogeneity concerns, making it difficult to identify a stable causal estimand.

Second, the temporal dynamics of social media use further complicate scientific inference. The potential effects of social media engagement on mental health and well-being are cumulative processes that unfold over time [59]. Detecting meaningful effects therefore requires modeling trajectories rather than relying on cross-sectional snapshots [17,60,61]. However, longitudinal designs are resource-intensive and analytically complex. Moreover, short-term and long-term effects may diverge. Consistent with theories of hyperbolic discounting [62,63], social media engagement may provide short-term gratification through news feeds and social interaction, while potentially conflicting with longer-term goals or aspirations [7,64].

Third, limited access to users' objective social media behaviors constitutes another structural constraint in the study of digital well-being. Early research relied primarily on self-

reported time spent that is vulnerable to recall bias [65,66]. Researchers subsequently sought objective behavioral data through platform Application Programming Interfaces (APIs), passive monitoring tools, and data donation packages that allow participants to share account histories [67,68]. However, access to high-quality objective data remains constrained. Establishing infrastructure that enables platforms or users to share data with independent researchers has proven difficult due to privacy and safety concerns, technical limitations, and limited incentives [69]. Moreover, even when objective behavioral data are obtained, researchers often face a methodological tension between high-dimensional digital traces and relatively modest sample sizes, where granular behavioral features can outpace statistical power and theoretical interpretability. Unlike fields that rely on large-scale population or administrative datasets with standardized, low-dimensional measures, digital well-being research must navigate a trade-off between measurement richness and conceptual clarity. Some progress has emerged: the European Union's Digital Services Act (DSA) strengthens users' rights to access their data [24]; academic initiatives are developing large-scale data donation systems [70,71]; and some platforms have begun collaborating with researchers [72], yet the sustainability and scalability of the collaborations remain uncertain.

Fourth, analyzing high-dimensional behavioral and content data of social media use remains challenging in digital well-being research. Although more precise measurement is expected to clarify the mechanisms linking social media use and well-being, extracting meaningful signals from complex data is difficult. For example, the Screenomics project captures screenshots of users' smartphones every five seconds while the screen is active, producing one of the most granular forms of behavioral data currently available [73]. As illustrated in their examples, sequential screenshots may include different platforms (e.g., Instagram, WhatsApp), distinct content types (e.g., cosmetic advertisements, transportation schedules), and varying interactions [74]. Systematically coding and interpreting such granular data is conceptually challenging and

analytically demanding. This challenge is similarly reflected in our analysis of objective TikTok usage in Chapter 3 of this dissertation.

A fifth methodological challenge in digital well-being research concerns the tension between Type I error control and hypothesis generation in high-dimensional exploratory analyses. Studies in this domain frequently examine a large number of exposure–outcome combinations, such as across platforms, app categories, behavioral metrics, and well-being dimensions, as illustrated by the review in Section 1.2 and the study in Chapter 2. Checking a large amount of exposure–outcome combinations creates a dilemma. On one hand, testing a large number of hypotheses simultaneously inflates the risk of false positive findings, making some form of correction for multiple comparisons necessary. On the other hand, many studies in this domain are better understood as hypothesis-generating than hypothesis-confirming. Specifically, as illustrated in Chapter 2, their goal is to identify candidate associations worthy of further investigation, not to test any single relationship. Applying strict corrections in this context substantially increases the risk of false negatives, such that associations with small but potentially meaningful effect sizes may be dismissed as evidence of absence. There is no consensus on how best to balance these competing demands. Increasing statistical power through larger samples and pre-registration represent sound practices. However, studies of complex digital engagement behaviors are resource-intensive, and achieving adequate power to detect small within-person effects reliably is not always feasible. Interpreting null results in high-dimensional exploratory analyses therefore requires careful distinction between "no effect" and "insufficient power to confirm a preliminary signal." The distinction here is analytically non-trivial and remains underappreciated in the literature.

1.4 Overview of the empirical chapters

The empirical chapters (Chapters 2–4) seek to advance understanding of digital well-being by addressing several of the methodological and conceptual challenges outlined above. Chapter

2 advances the measurement of social media use by introducing a scalable approach, using user-shared screenshots of device screen time dashboards, to collect objective, app-specific behavioral data. Drawing on these app-level metrics, Chapter 2 examines how platform-specific usage patterns relate to well-being outcomes within an EMA study of adolescents. Chapter 3 further extends objective measurement by incorporating TikTok data donation packages alongside screenshot-based app metrics. More importantly, Chapter 3 moves beyond identifying specific types of use as risk factors to investigate a broader psychological process that may underline different types of social media use. Chapter 3 conceptualizes social media engagement as a series of momentary goal pursuit episodes and examines whether goal awareness and goal strength predict subsequent experiences and well-being. Finally, Chapter 4 examines digital well-being practices from the platform perspective. While policymakers and scholars frequently debate platform responsibility, systematic documentation of platform-initiated well-being efforts remains limited. In Chapter 4, I conduct a large-scale content analysis of public announcements from seven widely used social media platforms, analyzing the claims, design features, target harms, and evaluation plans articulated by the platforms themselves. Together, these chapters advance research on social media and well-being at three levels: improving behavioral measurement, identifying cross-cutting psychological processes, and documenting institutional practices.

Chapter 2: Appnome analysis reveals small or no associations between social media app-specific usage and adolescent well-being

Authors

Yuning Liu¹, Laura Marciano²

Affiliations

¹ Harvard Kenneth C. Griffin Graduate School of Arts and Sciences, Harvard T.H. Chan School of Public Health

² The Media School, Indiana University Bloomington

Abstract

The debate on how social media use influences adolescent well-being is mostly based on self-reports of social media use (SMU). By collecting data and screenshots donated from 374 Swiss adolescents (Mean age= 15.71; SD age=0.82) over 2 weeks, we created “Appnomes” – app-specific usage metrics on screentime, number of activations, and number of notifications per participant per day, and associated them with daily hedonic and eudaimonic well-being. Longer TikTok time predicted lower eudaimonic well-being (standardized coefficient $\beta = -0.08$) but higher positive emotions ($\beta = 0.06$) the next day. In practical terms, a one standard deviation (SD) increases in TikTok use (approximately 39 minutes above an individual’s typical daily use) was associated with a 0.08 SD decrease in eudaimonic well-being the same day and 0.06 standard deviation increase in positive emotions the next day. Longer use of WhatsApp predicted negative emotions ($\beta = 0.06$, one SD=30 minutes above an individual’s typical use) while more screen activations for WhatsApp predicted greater feelings of connection ($\beta = 0.08$, one SD=17 activations above an individual’s typical levels). Instagram notification was positively related to increased feeling of focus ($\beta = 0.06$, one SD=12 notifications above an individual’s

typical levels) the next day. YouTube screen unlocks predicted more feelings of meaning ($\beta=0.07$, one SD=2 activations above an individual's typical levels) the next day. More Snapchat notifications predicted less relaxed, less competent, and less positive emotions (with $-0.07 < \beta < -0.06$, one SD=8 notifications above an individual's typical levels) the next day. Results pointed towards minimal or no effects, challenging the moral panic on the detrimental impact of social media use on teens well-being.

2.1 Introduction

In the recent decade, adolescents represent the fastest-growing demographic of social media users. In 2022, the time spent online by U.S. adolescents doubled, compared to 2014-15, with 95 % of teens using YouTube, 67% TikTok, 62% Instagram, and 59% Snapchat [75]. A similar trend can be found in Switzerland, where 99% of 12 to 19-year-olds own a smartphone, 95% use instant messaging apps, and 91% use social media sites, with Instagram, YouTube, Snapchat, and TikTok being the top ones [76]. Along with the increased use of social media, mental health problems are also rising among adolescents. Studies found that life satisfaction is lowest among individuals aged 15-24 years old in most of the world's regions [77]. In 2021, almost half of U.S. teens felt persistently sad or hopeless, and one-third reported poor mental health [78], while about 37% of Swiss teens experienced mental health problems [79]. Moreover, adolescence is the time window of developmental sensitivity to the negative well-being outcomes of social media [59]. As stated in the U.S. Surgeon General Advisory on Social Media and Youth Mental Health [17] and the update of the Digital Services Act in early 2024 [24], parallel trends of rise in adolescents' mental health problems and social media use ring an alarm bell for our society. Hence, exploring the impact of social media and smartphone use on adolescent well-being – with objective social media data - is increasingly vital to provide robust results beyond cross-sectional data and self-reported measures, and inform policymaking.

Despite some compelling studies [20,80,81], the (bi)-directional relationship between social media use (SMU) and adolescent well-being remains debated. In the release of the U.S. Surgeon General stated that “there are ample indicators that social media can also have a profound risk of harm to the mental health and well-being of children and adolescents” (p. 4) but called for more research using advanced methodologies to disentangle both negative and (understudied) positive effects of social media use on adolescents’ well-being [17]. So far, the literature on social media use and adolescent well-being includes many systematic reviews and reviews of reviews [27,28,82], and an emerging trend of intensive longitudinal studies using Ecological Momentary Assessments (EMAs) [54,83–85], and, in some rare cases, trace data [85,86]. The consensus among previous studies is that the extent of social media use has a negative but very small impact on adolescents’ mental health [27,28], but mechanisms behind the association need to be further examined.

In the sections below, we present a review of the four perspectives for advancing studies on social media use and adolescent well-being and discuss how the current study will incorporate these perspectives.

Current objective social media data collection

Central to the debate between social media use and well-being is identifying how activities on social media affect well-being, a task complicated by the idiosyncratic nature of social media use behavior. Current self-reported measures of social media time are inadequate as they fail to capture different activities and are often underreported [87] or distorted by users [88], and are only moderately correlated with objective metrics [89]. Therefore, granular and accurate data on users’ social media activity are essential to unpack the detailed mechanisms between social media use and adolescent well-being. The European Union’s DSA, effective February 2024, is a significant advancement that mandates platforms to share data on platform design and operation through appropriate interfaces with researchers [90] in future studies. Although some US states

have pilot programs [91], similar comprehensive legislation is lacking in other global regions, making collecting objective social media data for empirical research challenging.

Objective social media data are activity records as processed or saved through an online environment [92] and can be collected using APIs, app data donation, and passive sensing, but each of these approaches is often hampered by various challenges [67]. First, data from APIs are constrained by content limitations to protect user privacy, removing most identifiable information, such as the user demographics. This leads to the second limitation: it is difficult to link social media use data requested from API to well-being and mental health outcomes, which are mostly measured through self-reports [67,93]. Third, data from API are subject to rate limits and platform policy changes. For example, the Twitter Academic API and Reddit Pushshift API, once free to access for academic research, were both gated with subscription fees in 2023 [94,95].

Another option is using Data Donation Packages, archives of social media platforms that each user has the right to download. In both Europe and the United States, social media platforms are legally mandated to share user trace data with their users. However, challenges remain at both the data collection and analysis stages, and privacy issues. A low compliance rate has been reported for these kind of studies [68], and analyzing the content of data donation packages is often complicated by the frequent changes in infrastructure implemented by social media platforms [68].

Passive sensing apps like EARS and Ethica (now Avicenna) collect data on screen state and battery status from smartphones without the user's active input [96]. However, most sensing apps only collect overall screen time, and only a few Android-specific sensing apps can track the exact screen on and off times by app [97]. To partially overcome these limitations, the Human Screenome Project proposes Screenomic analysis, capturing smartphone screenshots every 5 seconds while the user is actively using the device [74]. While this method offers detailed data, this approach can be highly intrusive and increases privacy risks [98], especially for underage users. In addition, high-frequency data collection generates a vast amount of image data,

requiring substantial computational effort, and posing challenges in deriving meaningful measurements from the images.

In summary, collecting objective and specific data on teen smartphone usage and social media use at scale, while adhering to international regulations and minimizing privacy invasion among underage populations, presents significant research challenges. The current study introduces a novel method for collecting objective social media data from adolescents using user-donated screenshots to track specific app usages on smartphones. Reflecting the concept of screenomes, which records sequences of digital screen states, we developed "Appnomes"—metrics that measure app-specific screen time, number of activations, and number of notifications per participant per day, derived from user-donated screenshots. Appnomes encapsulate the body of app-specific usage information collected in this manner. For more detailed information on the methodology for collecting and analyzing user-donated screenshots, please refer to a previous study by our team [99].

Between- and within-level analysis in social media use and well-being

The effects of social media use on adolescent well-being are person-specific and highly heterogeneous [81], necessitating both between- and within-level analyses to capture the underlying complexity [20]. Intensive longitudinal designs such as EMAs enable multilevel modeling, where each survey assessment (level one) is nested within participants (level two) [61]. In multilevel modeling, the between-level analysis examines intergroup relationships, while within-level analysis focuses on intragroup levels, examining how social media use predictors at different times are associated with the well-being of the same individual [100] (e.g., N=1 approach [20]).

Multilevel modeling in social media use and well-being studies enable researchers to compare how between- and within-person level associations differ. Beyens et al. [81] found no significant average within-person effects between active and passive social media use and well-being, yet observed substantial heterogeneity with some adolescents experiencing positive

associations between active use and well-being effects and others experiencing negative associations. Moreover, the association varied among adolescents with 44% feeling no change, 46% feeling better, and 10% feeling worse after passive SMU. Orben et al. [101] analyzed data from seven waves of a UK national survey, noting small negative between-level associations between social media use and life satisfaction, but not at the within-person level. Other studies have reported differences in the between- and within-level analysis results [54,85,102,103]. A key challenge in this approach is its inability to support inferences from an individual to a targeted population. Therefore, scholars advocated for principles to quantify, interpret, and explain variation in average within-person effects [104]. These findings indicate that simply focusing on between-level analyses can obscure important individual dynamics. Therefore, both within- and between-level analyses, combined with longitudinal designs, should be conducted in studying social media use and well-being.

Temporal dynamics of social media use and well-being

Recent discussion in research has emphasized the temporal dynamics of the association between social media use and well-being, exploring the "delayed effects" where social media use at one time point relates to well-being later. For example, a study [105] with 89 young adults found that passive Facebook usage was related to reduced lagged affective well-being two hours later. Another experience sampling study found that social media use was positively associated with perceived uncertainty at the same measurement occasion, but not at the subsequent measurement occasion, suggesting a short-lived relationship [106], suggesting a transient relationship. Marciano et al.[85] found that frequent smartphone use significantly predicted improved well-being the day after, and this effect size shrank at three days and returned to near zero at seven days. These studies underscore the temporal dynamics, suggesting that the potential of social media use on well-being are transient.

However, determining the appropriate time unit for longitudinal studies analyzing the social media use and well-being relationship—minutes, hours, days, or months—remains challenging. While exposure to social media can trigger neural reactions within seconds [107], most EMA studies collect data from the previous hours or the current day, and cohort studies, like those by Orben et al. [101], aggregate associations over even longer periods. Indeed, panel studies that track participants for months or years may not account for these temporal dynamics, potentially biasing study interpretations. This mismatch in time intervals, between the rapid neural responses and the feasible measurement intervals in observational studies, leaves unresolved questions (e.g., how effects can accumulate and interact through time?). Ephemeral associations (effects lasting only a few minutes) may not be captured, while longer-term associations (effects over several years) remain unexamined. More empirical data is needed to identify these dynamics across different time scales. In the current study, we examine how social media use is associated with well-being on the same day and with one- and two-day lags.

App-specific social media components and psychological processes

In examining social media use and well-being, it is essential to recognize that different apps offer unique features, functions, and content, which result in different user activities and experience. Studies have distinguished between active usage, involving user interactions, and passive usage, which focuses on content consumption on social media [18]. Recent research has identified core components of the social media ecosystem and the psychological processes related to them, such as profiling with self-presentation, streaming that triggers social comparison, and networking with receiving social support [108].

Different social media apps emphasize these components to varying degrees. For instance, WhatsApp mainly provides channels for networking and messaging, TikTok is more on providing streaming through short videos, while Facebook and Instagram blend these components. This variation in emphasis leads to differing psychological processes, and

subsequently, distinct relationships between platform use and well-being across different apps. For example, during COVID-19, a cross-sectional study found that passive Facebook use was linked to decreased well-being through social comparison, active Instagram use was associated with enhanced life satisfaction via social support, while TikTok had no significant impact [109]. Another EMA study has shown that both Facebook and Instagram use predicted increased social comparison, lower self-esteem, and negative affect [110]. Additionally, a two-wave panel survey found perceived information overload in YouTube among adults, in WhatsApp and Snapchat among older adults, but not in Facebook or Instagram [111]. Another experimental study found no significant differences in subjective well-being between active and passive uses of Instagram, Facebook, or Twitter [112]. Most of these studies targeted adults, yet there is evidence that adolescents use Facebook less and prefer TikTok and Instagram more than adults [75]. Together, these studies emphasize the importance of moving beyond general metrics of social media use to focus on specific apps, especially among adolescents. Moreover, rather than merely focusing on specific social media platforms, studying the app-specific use with well-being facilitates identifying unique platform features and underlying psychological processes, which further supports the study of psychological mechanisms between social media use and well-being.

Our research also explores the relationship between gaming apps like "Clash of Clans" and "Pokemon Go," and well-being, that are less studied than social media apps in existing literature. Previous studies have undermined the idea that video gaming harms mental health, showing minimal evidence of harm and, in some cases, noting benefits such as stress reduction [113–115]. Digital games may meet various psychological needs, such as identification, social interaction, and escapism, linked to improved hedonic, eudaimonic, and social well-being [115]. Yet, research also shows a significant negative impact of problematic gaming on adolescents' mental and psychosocial health [116,117]. Given that adolescents spend as much time gaming as they do using social media [118], and both activities engage similar psychological processes,

there is a strong case for comparing the impact of gaming apps and social media apps on well-being.

From ill-being to well-being

Another important perspective for studies on social media use is to move beyond focusing on ill-being, such as psychopathological symptoms, and shift towards well-being and flourishing [119]. This shift is important because the absence of ill-being does not necessarily equate to well-being [120]. Moreover, this transition reflects a broader philosophical change in social media studies: moving from identifying failures of social media platforms to exploring how platform designs or user behaviors can promote well-being [30,121].

Responding to the call for a broader perspective, our study adopts the human flourishing framework to examine well-being as a cornerstone for societal thriving [122]. We focus on hedonic well-being, happiness or contentment derived from experiencing pleasure and avoiding pain, and eudaimonic well-being [122], rooted in Aristotle's concept of self-actualization and meaningful life purposes. Hedonic well-being includes aspects such as life satisfaction, positive affect, and negative affect [123]. Theoretical dimensions of eudaimonic well-being include autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance [123]. A study showed how specific social media activities are related to adolescent flourishing levels, however, only between-person associations have been explored [119]. A recent meta-analysis summarized the association between social media use and "positive" well-being indicators finding that hedonic well-being was positively related to communication activities on social media (Pearson correlation coefficient, $r=0.11$) and positive online experiences ($r=0.21$), problematic social media use ($r=-0.13$) and social comparison on social media ($r=-0.30$). While eudaimonic well-being was only associated with problematic social media use ($r=-0.26$) [30]. However, included studies were mostly cross-sectional and none used intensive longitudinal data with social media data collected through screenshots.

The present study

The present study provides a possible solution to address previous issues on collecting objective social media data to disentangle the between- and within-person variation, examine the temporal dynamics, and consider adolescent well-being outcomes in a continuum from ill- to well-being. Using Appnoms—app usage metrics on screentime, number of activations, and number of notifications derived from user-donated screenshots—and well-being measures from EMAs, we applied multilevel modeling to explore between- and within-person associations and answer the following two research questions: RQ1: How do Appnoms relate to well-being at the between- and within-person level? RQ2: Is the relationship between App-specific usage metrics and well-being consistent over time?

The HappyB study applied an EMA design and collected data on the well-being of 374 adolescents living in Switzerland ($\text{Mean}_{\text{age}} = 15.71$; $\text{SD}_{\text{age}} = 0.82$, 62.8% female) over two weeks in Spring 2022. Participants were invited to rate their well-being three times per day (with prompts at 12 pm, 6 pm, and 9 pm) with questions related to the last two hours. In total, 277 participants completed 2,273 assessments and uploaded screenshots, achieving a 58.6% compliance rate at the assessment level. Well-being measures included positive emotions (i.e., happiness, relaxation, and contentment) and negative emotions (i.e., anger, sadness, and loneliness), while eudaimonic well-being was related to the engagement in the activities (i.e., feeling “engaged”, “competent”, and “focused”, and feeling that the activities “were worthwhile and meaningful”, “made me feel connected with others”, and “made me feel satisfied”). Answer options were on a Visual Analogue Scale (VAS) from 0 “not at all” to 100 “A lot”. Participants also donated daily screenshots of App-specific usage metrics, including the total screentime, number of unlocks, and number of notifications, for the three most used apps from the previous days. The study received ethical approval from the IRB of USI, Università della Svizzera Italiana, Lugano, Switzerland, and was supported by the Department of Education, Culture, and Sport of Canton Ticino. Informed

consent was obtained from all study participants following regulations of the involved institutes in Switzerland. All data collection and study analyses were performed in accordance with relevant guidelines and regulations of the involved institutes.

2.2 Method

Data Collection

The HappyB study included the collection of intensive longitudinal data by tracking social media use and well-being in 374 Swiss adolescents (Mean-age= 15.71; SD-age= 0.82, 62.8% female) over 14 days in Spring 2022. Note that the target sample size for our EMA was set at 250 participants prior data collection, based on power analysis to achieve 90% power for detecting small effect sizes (0.1 SD). Students who were in grades 1 and 2 are recruited from 4 high schools in Switzerland and joined the study between the 7th and the 23rd of March, 2022. Participants completed a baseline survey and EMAs using the Ethica app (now Avicenna). The timeline of the HappyB study data collection is shown in Figure A1.1 in the supplementary file. The baseline survey collected information on the socio-demographic characteristics, while EMAs were conducted three times daily from Day 1 to Day 14 at 12 pm, 6 pm, and 9 pm via the Ethica App. EMAs measured self-reported current well-being including hedonic and eudaimonic well-being assessment. Participants also provided screenshots of phones' settings, showing App-specific usage patterns including daily usage time, screen activations, and notification counts of used apps. The example screenshots are shown in Figure A1.2. To collect the objective data on social media App-specific usage patterns, we ask participants to go to their smartphone's Settings → Usage Time, take screenshots of platform usage information from yesterday, and upload it to Ethica. The platform usage information includes usage time, number of screen unlocks, and number of notifications.

Well-being measurements

On a scale of 0-100, participants were asked to report the level of positive emotion (measured by happiness, relaxation, and contentment), negative emotion (measured by anger, sadness, and loneliness), eudaimonic activity (measured by the feeling engaged, competent, and focused), and eudaimonic well-being (measured by feeling meaningful, connected, and satisfied) in the previous two hours. The well-being measures were collected three times per day in the study period and averaged into daily measures in the downstream analyses. The distributions of specific and aggregated well-being measures are presented in Supplementary Figure A1.3 and Supplementary Figure A1.4.

Text extraction from user-donated screenshots

We extract information on App-specific usage patterns, including the total and app-specific screentime, number of notifications, and number of activations, from the collected screenshots. For example, in Supplementary Figure A1.2 (A), we extract the following information for an individual: a total screentime of 278 minutes (4h 38m) from the previous day, with 63 minutes from Chrome, 53 minutes from Office, and 47 minutes from YouTube. Similarly, in Supplementary Figure A1.2 (B), we extracted: a total of 157 screen activations happened the previous day, with 44 activations due to WhatsApp, 19 activations for Mobile FFS and another 19 activations for Photomath.

We developed a pipeline to extract details on total screen time, number of activations, and notifications for the top 3 apps from screenshots. The text extraction pipeline incorporates image preprocessing, Tesseract Optical character recognition (OCR), text translation, and post-processing [99]. OCR is a method to convert texts from images into machine-encoded text. We applied OCR to transfer the objective social media usage data from the screenshots to text. Then, we conducted post hoc heuristic correction of the extracted texts, correcting the extracted app names based on a database of app names using edit distance.

To evaluate the accuracy of our text extraction, we randomly selected a total of 600 screenshots, with 100 each for iOS/Android screen time, activation, and notification, and annotated them manually to establish the ground truth. For example, the ground-truth labels for Figure A1.2 (C) are: 347, “WhatsApp”, “267”, “TikTok”, “24”, “Instagram”, and “17”. In general, we generated 7 measures for each screenshot, including the total screentime/count, app name for the app with the 1st screentime/count, screentime/count for the 1st app, app name for the app with the 2nd screentime/count, screentime/count for the 2nd app, app name for the app with the 3rd screentime/count, and screentime/count for the 3rd app. We evaluate the pipeline's performance based on the accuracy rate—the percentage of screenshots where extracted text precisely matched the ground truth, with the results shown in Supplementary Table A1.1.

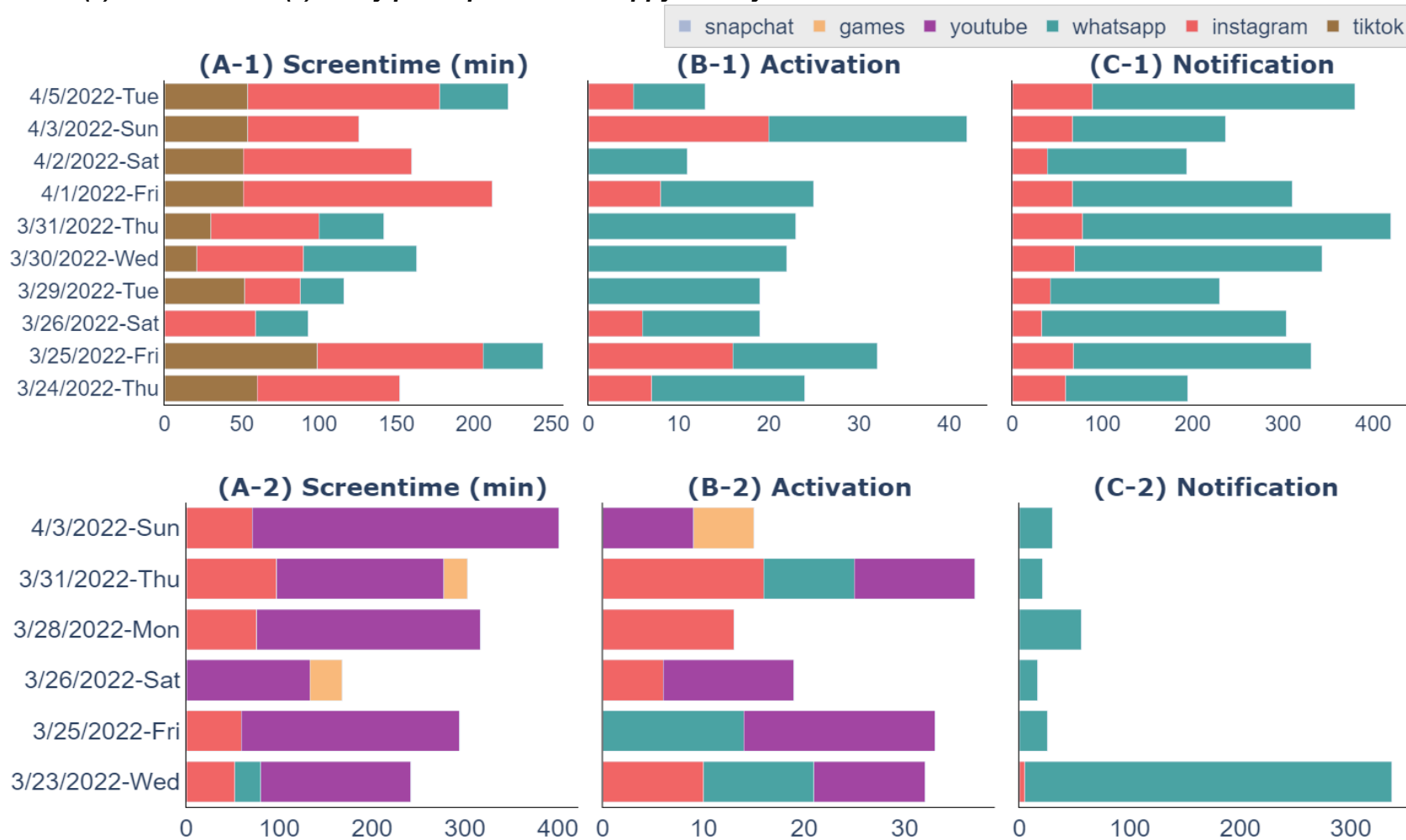
Appnome construction

To establish App-specific usage patterns for each participant, we generated an Appnome that displays daily screen time, activation count, and notifications per app and app category across the study period. In this study, we constructed the Appnome by the top 6 Apps with the highest average rank in screen time, number of activations, and number of notifications, including TikTok, WhatsApp, Instagram, YouTube, Game apps, and Snapchat, in the downstream analyses. Note that we aggregated all game apps into the general Games category in the app-specific analysis. Using the text extracted from screenshots, we created an Appnome for each participant, detailing daily screen time, notifications, and activations for the top three apps. This is visualized for each participant, as illustrated in Figure 1.1. We also describe the distribution of person-level average screentime, number of activations, and number of notifications in Figure 1.2.

Before downstream analyses, we applied a series of filters to ensure the data quality. We preprocessed the text extracted from the screenshots at both the screenshot and participant level. At the screenshot level, we dropped screenshots without total counts, removed screenshots without the name of the first app, excluded screentime screenshots with total screen time or app-

specific screentime exceeding 20 hours (1200 minutes), dropped screenshots with the total screentime (or counts) smaller than the sum of the screentime (or count) of the top three apps, and removed screenshots with the screentime (or counts) of the first app smaller than the screentime (or counts) of the second app as well as the third app, as these scenarios could indicate a measurement error, such as the participant submitting a weekly instead of a daily view, or an error with the text extraction pipeline. At the participants level, we kept participants that had at least 2 days of observations with both the App-specific usage patterns and the well-being outcome measures. In the filtered sample, we have 1,151 assessments from 151 people for screentime analysis, 1,105 assessments from 150 people for activation analysis, and 1,357 assessments from 191 people for the notification analysis.

Figure 1.1 An example of Appnome on the screentime (A), number of activations (B), and number of notifications (C), of one female (1) and one male (2) study participant in the HappyB study



Note: The figure shows the App-use features of one female adolescent and one male adolescent in the study. Study participants uploaded daily screenshots from their cellphone settings, detailing App-specific screen time, unlocks, and notifications throughout the study period. Image processing and OCR techniques were employed to extract text from these screenshots.

-- As shown in (A-1), the female adolescent spent most of the screen time on TikTok and Instagram in almost the entire study period. As shown in (B-1), most of the screen unlocks comes from activating WhatsApp. As shown in (C), the adolescent received more than 150 notifications from WhatsApp daily and less than 100 notifications from Instagram daily in the study period.

-- The male adolescent has a different app-use feature compared to the female participant. As shown in (A-2), the male adolescent spent more than 1000 minutes using YouTube in the observed days. The adolescent also spent some time on Instagram and Game Apps. As shown in (B-2), most of the screen unlocks comes from activating Instagram, WhatsApp, and YouTube. As shown in (C-2), the adolescent received notifications mainly from WhatsApp in the study period.

Between and within-level analyses on app usage and well-being

First, we conducted a between-level analysis by calculating partial correlations between the average well-being score of the participants and the average app usage across study periods, adjusting for age and gender (Figure 1.3). The Bonferroni correction is conducted to correct for multiple hypotheses testing. For example, regarding positive emotion with app-specific screentime, six regression models are fitted to test competing hypotheses, thus the level of confidence is $0.05/6 = 0.0083$.

Then, we analyzed well-being outcomes (positive and negative emotions, eudaimonic activity, and eudaimonic well-being) and app usage features at level one (days), nested within individuals at level two, with the lme4 package in R (4.2.1) software [124]. Using mixed linear models, we explored within-person variations in well-being outcomes (Y) and App-specific features (X) of the same day, adjusting for well-being outcome of the previous day (Y_{t-1}), age, gender, and school, with days as level one and individuals as level two. For example, positive emotion (y_1) -- screentime of Instagram (x_1) is one pair of well-being outcomes – App-use feature predictors. In total, we have 288 pairs of such outcomes (12 outcomes and 4 aggregated outcomes) and predictors (3 features per app for 6 Apps), that we fit mixed linear models on. For each well-being outcome and App-use feature, we adjust the P values by $P_{\text{adjusted}} = P_{\text{unadjusted}} * 6$ to control for multiple hypotheses testing. The coefficient estimates that are significant after the Bonferroni correction are marked with * and highlighted in bold.

For each outcome and feature pair, we fitted two linear regression models, one with a random intercept, and one with a random intercept and a random slope. We then conducted likelihood ratio tests comparing the two models to select the model that provides a good fit and is more parsimonious. The App-specific usage features were person-mean centered in the mixed linear models, with the random intercept models specified as:

$$\begin{aligned} Wellbeing_{ti} = & \beta_{0i} + \beta_1 AppFeature_{ti} + \beta_2 Wellbeing_{(t-1)i} + \beta_3 Age_i + \beta_4 Gender_i + \beta_5 HighSchool_i \\ & + e_{ti} \end{aligned}$$

$$\beta_{0i} = \gamma_{00} + \mu_{0i}$$

And the random slope random intercept model is specified as:

$$\begin{aligned} Wellbeing_{ti} = & \beta_{0i} + \beta_{1i}AppFeature_{ti} + \beta_{2i}Wellbeing_{(t-1)i} + \beta_{3i}Age_i + \beta_{4i}Gender_i \\ & + \beta_{5i}HighSchool_i + e_{ti} \end{aligned}$$

$$\beta_{0i} = \gamma_{00} + \mu_{0i}$$

$$\beta_{1i} = \gamma_{10} + \mu_{1i}$$

The mixed linear model results are presented in Table 1.1, zooming in on the relationship between the four aggregated well-being outcomes, positive/negative emotion and eudaimonic activity/well-being, and App-specific usage features. In addition, the correlation between App-specific usage features and individual well-being measures, such as sadness and feeling satisfied, are presented in Figure 1.4 (presenting results for TikTok, WhatsApp, and Instagram) and Supplementary Figure A1.5 (presenting results for Games Apps, YouTube, and Snapchat).

Same-day and 1-, and 2- day lagged within-level analyses on app usage and well-being

Next, we extended the same-day analysis to 1- and 2-day lagged analyses to test intra-individual relationships between well-being outcomes and App-specific usage features using multilevel modeling. Similar to the same-day analysis, we conducted Bonferroni correction for multiple testing. The 1-day lag random intercept model is specified as:

$$\begin{aligned} Wellbeing_{ti} = & \beta_{0i} + \beta_{1i}AppFeture_{(t-1)i} + \beta_{2i}Wellbeing_{(t-1)i} + \beta_{3i}Age_i + \beta_{4i}Gender_i \\ & + \beta_{5i}HighSchool_i + e_{ti} \end{aligned}$$

$$\beta_{0i} = \gamma_{00} + \mu_{0i}$$

And the 1-day lag random slope random intercept model is specified as:

$$Wellbeing_{ti} = \beta_{0i} + \beta_{1i}AppFeature_{(t-1)i} + \beta_{2i}Wellbeing_{(t-1)i} + \beta_3Age_i + \beta_4Gender_i \\ + \beta_5HighSchool_i + e_{ti}$$

$$\beta_{0i} = \gamma_{00} + \mu_{0i}$$

$$\beta_{1i} = \gamma_{10} + \mu_{1i}$$

The results comparing the same-day, 1-day, and 2-day lagged multilevel models for TikTok, WhatsApp, and Instagram are presented in Figure 1.5, and the results for Games Apps, YouTube, and Snapchat are presented in Supplementary Figure A1.6.

Attrition analysis

We remove participants from data analysis if the participant completed less than 2 assessments across the entire study period. We conducted hypotheses testing to examine whether people dropped from the study have significant differences in demographic characteristics, App-specific usage features, and well-being outcomes. Analysis of Variance (ANOVA) and Chi-squared test were applied to check whether group means and frequencies were significantly different. The results of the attrition analysis are presented in Supplementary Table 3. Given the attrition we observed in our study, we also conducted a power analysis for the multilevel modeling using the mlmpower package in R[125]. Based on the observed data, the intraclass correlation coefficient (ICC) ranged from 0.45 to 0.6, and variance explained by the fixed effects of Level 1 and Level 2 predictors were around 0.003 and 0.0035, respectively. The estimated power results show that with $n_{\text{between}}=150$ and $n_{\text{within}}=7$, the power estimates for the app-specific usage variables ranges from 0.66 ± 0.07 to 0.75 ± 0.06 , and with $n_{\text{between}}=150$ and $n_{\text{within}}=10$, the power estimates ranges from 0.80 ± 0.08 to 0.88 ± 0.07 . We discussed the attrition issue in the Limitations section above. The data and code for replication can be found in the Open Science Framework (OSF) page [here](#).

2.3 Results

Individual “Appnome”

Using text extracted from user-donated screenshots, we constructed individual “Appnome”, summarizing total and top three apps’ metrics for each dimension (screentime, unlocks, and notifications) for each user across the study period. For the current analyses, we selected the top 6 social media Apps with the highest average rank in screen time, number of activations, and notifications for the analyses in this paper. Figure 1 (A-1, B-1, and C-1) shows the Appnome of one female participant who primarily used WhatsApp, TikTok, and Instagram. She spent 36-161 minutes on Instagram and 21-99 minutes on TikTok daily, unlocked her smartphone 8-23 times for WhatsApp, and received 136-342 WhatsApp notifications and around 50 Instagram notifications during the study. Figure 1 (A-2, B-2, and C-2) displays the Appnome of one male participant, who predominantly used WhatsApp, YouTube, and some Instagram. His daily YouTube screen time ranged from 130 to 330 minutes, with around 10 times of activations each for YouTube and Instagram, and 332 WhatsApp notifications on a Wednesday.

Between-person level app usage characteristics

Figure 1.2 and Table A1.2 show that females primarily used TikTok, averaging 78.19 minutes (SD=62.33 min) daily, followed by Instagram (mean=37.89 min, SD=38.13 min) and WhatsApp (mean=36.99 min, SD=33.92 min). Males used YouTube the most, averaging 44.27 minutes (SD=53.94 min) daily, followed by Instagram (mean=27.85 min, SD=30.38 min), WhatsApp (mean=25.11 min, SD=29.37 min) and TikTok (mean=24.93 min, SD=47.46 min). Females spent significantly more time on TikTok than males ($P<.001$), while males spent significantly more time on YouTube than females ($P=.001$). WhatsApp accounted for the most screen unlocks, with average unlocks of 33 (SD=26) for female and 23 (SD=23) for male daily. Both females (mean=114 times, SD=78 times) and males (mean=107 times, SD=113 times) received the most notifications from WhatsApp, followed by Instagram.

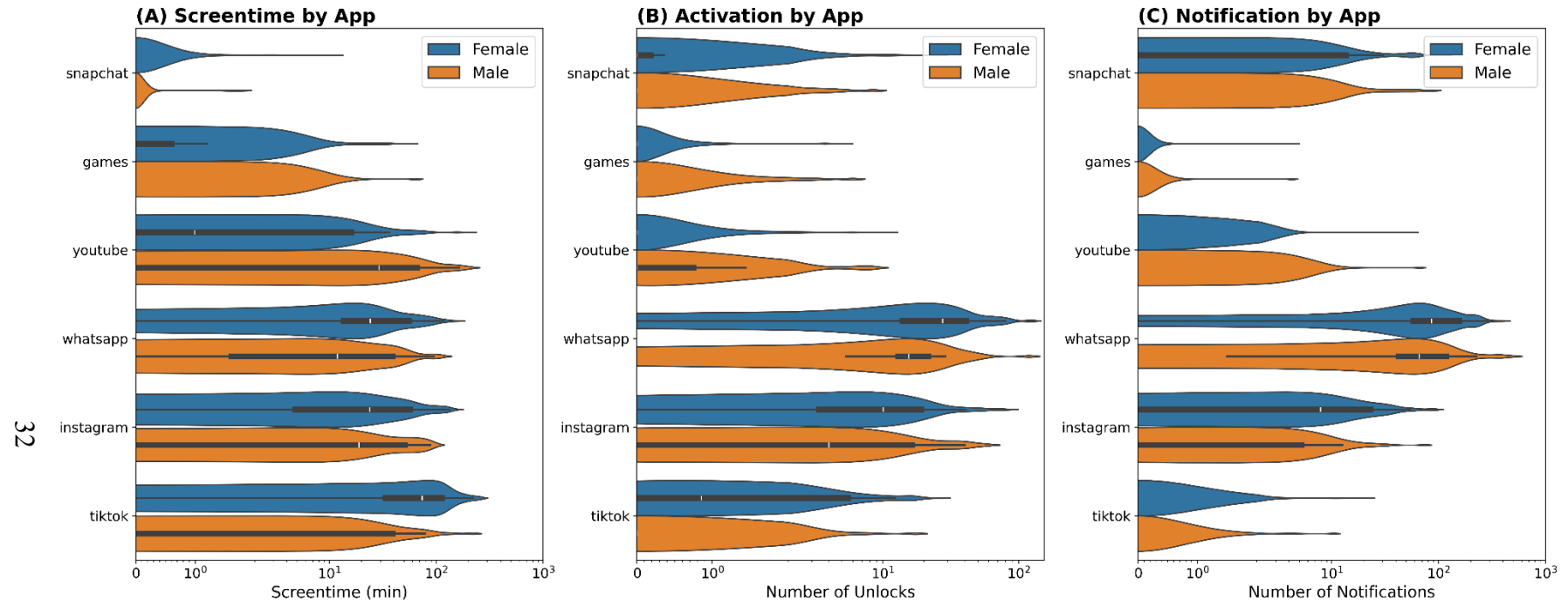
Between-person level correlation between app usage and well-being

We examined the partial correlation (controlled for age and gender) between the app-specific usage and the same day well-being outcomes at the between-person level. Figure 1.3 reveals that longer TikTok screentime ($r=-0.15$), more Games Apps activation ($r=-0.21$), and more Snapchat notifications ($r=-0.12$) correlated with lower positive emotion. Conversely, longer screen time ($r=-0.12$) and more notifications ($r=-0.12$) on YouTube were linked to lower negative emotions. None of the correlations were statistically significant following Bonferroni correction ($p<.008$).

Within-person level correlation between app usage and well-being

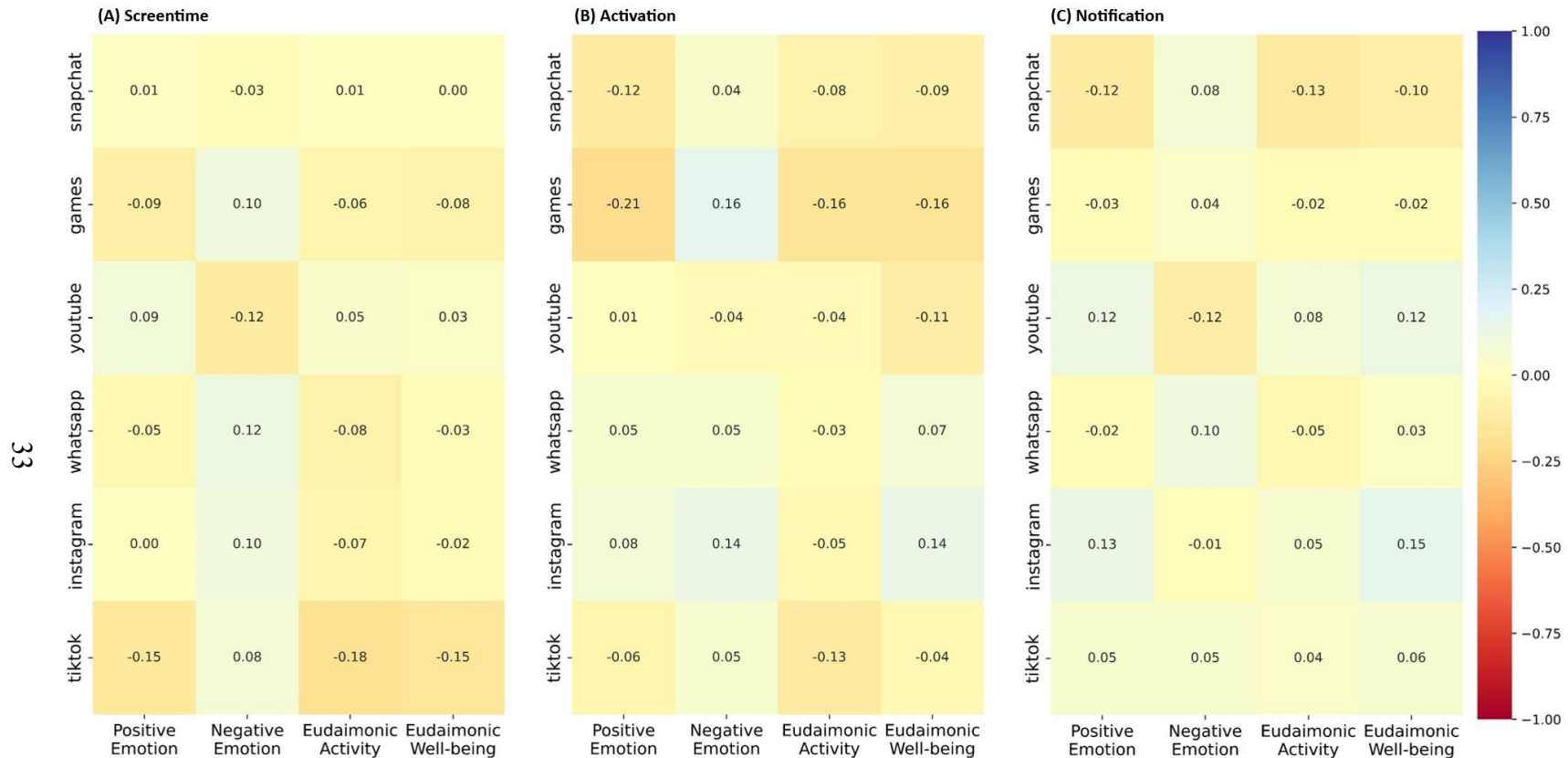
We conducted multilevel analyses, with time at level one and participant at level two, controlling for age, gender, school, and the previous day's well-being outcome (the autoregressive) to account for within-person variations, for each pair of app-specific usage and well-being outcomes. Table 1.1, Figure 1.4 and Figure A1.5 present the standardized and unstandardized correlation between app-specific features and well-being outcomes. As shown in Table 1.1, after Bonferroni correction, most correlations are not statistically significant. However, several features of apps like TikTok and WhatsApp show significant associations with well-being. For statistical significance, we set the threshold at $p<.008$ ($0.05/6$), accounting for 6 competing hypotheses for each pair of well-being outcomes and app features.

Figure 1.2. The distribution of person-level average screentime (A), number of activations (B), and number of notifications (C) by gender in the HappyB study



Note: The figure presents the distribution of the person-level average App-use features. Study participants uploaded daily screenshots from their cellphone settings, detailing app-specific screen time, unlocks, and notifications throughout the study period. OCR techniques were employed to extract text from these screenshots. Utilizing the extracted text, we calculated the average app-use features for each participant and depicted the distribution of these averages among the study participants in a violin plot.

Figure 1.3. Correlation between average screentime (A), number of activations (B), and number of notifications (C) of selected Apps and average hedonic and eudaimonic well-being in the between-person level in the HappyB study



Note: The figure presents the partial correlation of the between-person level well-being outcomes (Y) and App-use features (X) adjusting for age and gender. For instance, the average score of positive emotion (y1) and screentime on Snapchat (x1) are calculated for each person. The partial correlation coefficient between y1 and x1 is then estimated among the study participants. Bonferroni correction is conducted to correct for multiple testing (adjusted P value threshold = $0.05/\# = 0.0083$, where $\# = 6$). None of the correlation coefficients are significant statistically after the correction.

Table 1.1. Mixed linear model results on the standardized associations between screentime, number of activations, and number of notifications and hedonic/eudaimonic well-being in the HappyB study

	Positive Emotion				Negative Emotion				Eudaimonic Activities				Eudaimonic Well-being			
	Coef	SE	P	VPC	Coef	SE	P	VPC	Coef	SE	P	VPC	Coef	SE	P	VPC
Screentime by App																
Games	0.01	0.02	0.677	0.34	0.02	0.02	0.324	0.21	-0.04	0.02	0.043	0.46	-0.03	0.02	0.254	0.36
Instagram	0.00	0.02	0.980	0.33	0.02	0.03	0.549	1.00	-0.03	0.02	0.155	0.46	-0.02	0.02	0.423	0.36
Snapchat	0.01	0.02	0.712	0.33	0.02	0.02	0.495	0.20	-0.02	0.02	0.474	0.45	-0.02	0.02	0.417	0.35
TikTok	-0.04	0.02	0.044	0.33	0.03	0.02	0.258	0.21	-0.04	0.02	0.037	0.44	-0.08	0.02	0.001*	0.35
WhatsApp	-0.04	0.02	0.075	0.31	0.06	0.02	0.008*	0.19	0.00	0.02	0.882	0.45	-0.01	0.02	0.607	0.35
YouTube	0.03	0.02	0.178	0.33	-0.01	0.02	0.702	0.19	0.00	0.02	0.941	0.46	-0.03	0.02	0.206	0.37
Activation by App																
Games	0.00	0.03	0.907	0.17	0.00	0.03	0.903	0.99	-0.02	0.03	0.472	0.26	-0.04	0.03	0.108	0.21
Instagram	-0.01	0.03	0.763	0.16	0.02	0.02	0.503	0.37	-0.03	0.03	0.335	0.25	-0.01	0.03	0.817	0.22
Snapchat	0.02	0.04	0.587	0.98	0.01	0.02	0.840	0.38	0.01	0.03	0.596	0.25	0.02	0.03	0.528	0.21
TikTok	-0.03	0.03	0.238	0.17	0.01	0.02	0.712	0.39	-0.03	0.03	0.325	0.26	-0.02	0.03	0.429	0.21
WhatsApp	0.02	0.03	0.363	0.16	0.00	0.03	0.990	0.37	0.02	0.03	0.359	0.25	0.07	0.03	0.013	0.19
YouTube	0.03	0.03	0.211	0.17	-0.05	0.03	0.030	0.38	0.04	0.03	0.119	0.25	0.00	0.03	0.947	0.21
Notification by App																
Games	0.01	0.03	0.691	0.39	0.03	0.04	0.456	0.69	0.02	0.03	0.464	0.45	0.00	0.03	0.924	0.37
Instagram	0.04	0.02	0.053	0.37	-0.04	0.03	0.241	1.00	0.03	0.02	0.110	0.43	0.02	0.02	0.437	0.35
Snapchat	-0.01	0.02	0.486	0.34	0.01	0.02	0.750	0.25	-0.03	0.02	0.144	0.42	-0.04	0.02	0.105	0.35
TikTok	0.00	0.02	0.958	0.37	0.06	0.02	0.003*	0.26	0.00	0.02	0.827	0.43	-0.02	0.02	0.455	0.36
WhatsApp	-0.03	0.02	0.281	1.00	0.04	0.02	0.044	0.23	-0.03	0.02	0.103	0.42	-0.02	0.03	0.386	1.00
YouTube	-0.01	0.02	0.656	0.36	0.00	0.02	0.879	0.23	0.04	0.02	0.078	0.43	0.02	0.02	0.457	0.36

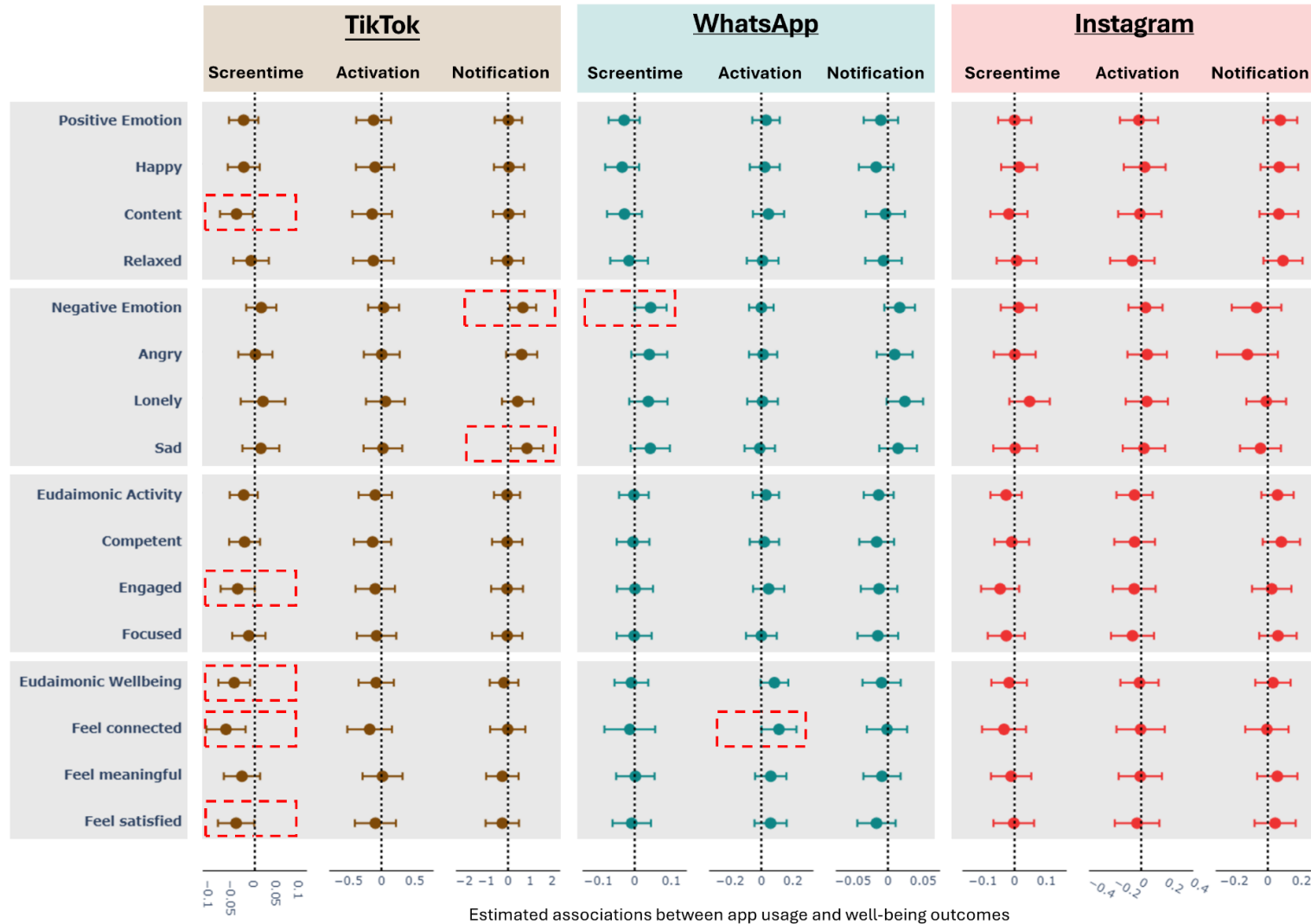
Note: The Table shows the standardized relationship between well-being outcomes (Y) and the App-specific features (X) of the same day from mixed linear models, adjusting for the well-being outcome of the previous day, age, gender, and school in each model, with time (each day) as level one, and adolescents as level two. Well-being outcomes are collected from self-reported questionnaires, and App-specific features are derived using the text extraction pipeline from the cellphone screenshots of the setting pages uploaded by users, both within the same EMA study.

-- For example, positive emotion (y1) -- screentime of Instagram (x1) is one pair of well-being outcomes – App-use feature predictors. We have 72 pairs of such outcomes (4 outcomes) and predictors (3 features per App for 6 Apps).

-- For each of the 72 pairs of well-being outcomes – App-use feature predictors, we fit a random intercept model and random slope model. We apply the likelihood ratio test to select the best model following the parsimonious rule. We report the estimated coefficients, standard errors, P values, and the Variance partitioning coefficients (VPC) of each model in Table 1.1.

-- **Bonferroni correction is applied to account for multiple testing.** For each well-being outcome and App-use feature, we apply an adjusted threshold for the P values. For example, regarding positive emotion with screentime, six regression models are fitted to test competing hypotheses, thus the P value threshold for statistical significance is adjusted by $0.05/6 = 0.0083$. The coefficient estimates that are significant after the Bonferroni correction are marked with * and highlighted in bold.

Figure 1.4. Forest plot of the mixed linear models results on the associations between TikTok, WhatsApp, and Instagram screentime, number of activations, and number of notifications and hedonic/eudaimonic well-being in the HappyB study



Note: The Figure shows the unstandardized coefficients and confidence interval between well-being outcomes (Y) and App-specific features (X) of the same day from mixed linear models, adjusting for well-being outcome of the previous day, age, gender, and school in each model, with time

(each day) as level one, and adolescents as level two. Well-being outcomes are collected from self-reported questionnaires, and App-specific features are derived using text extraction pipeline from the cellphone screenshots of the setting pages uploaded by users, both within the same EMA study.

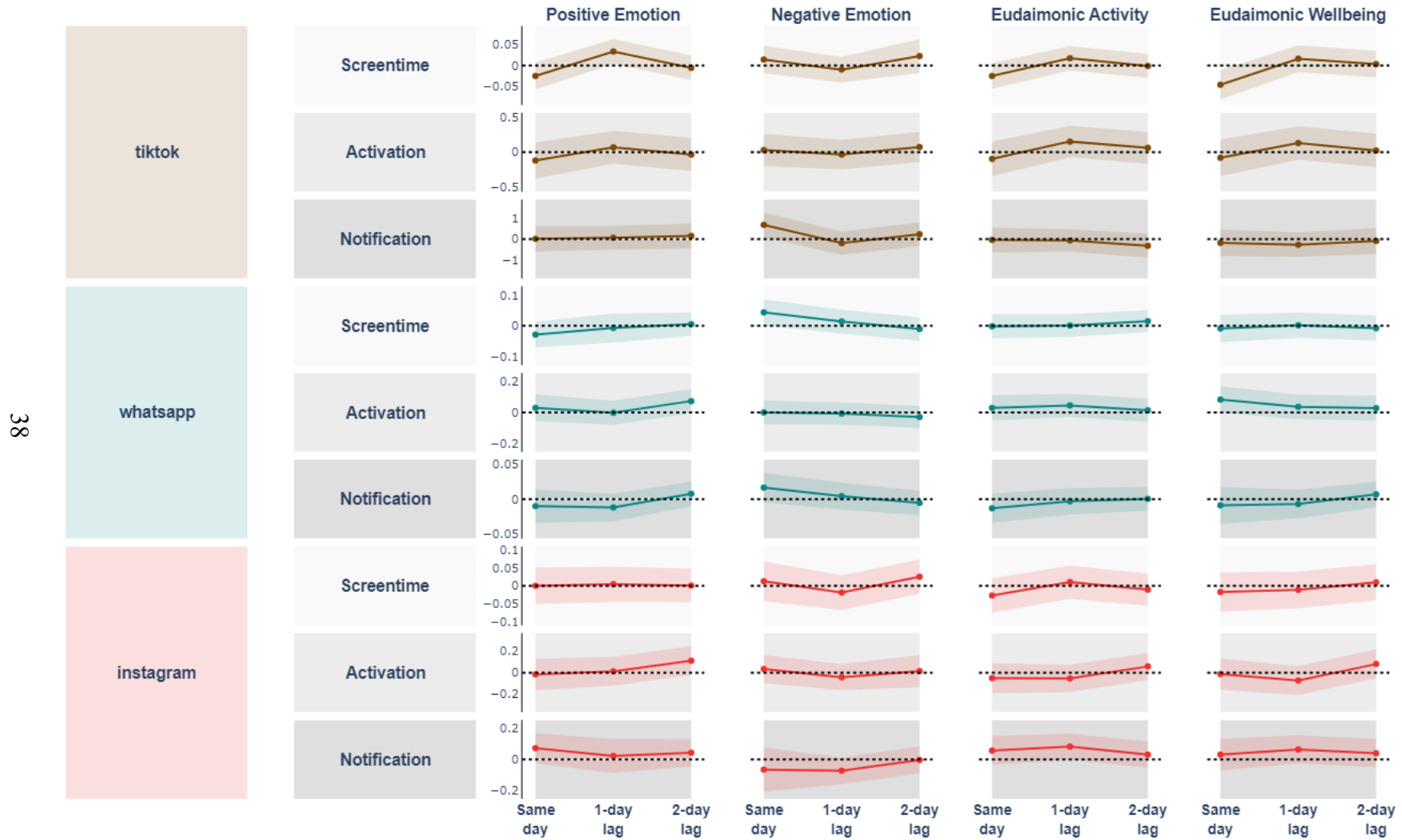
-- For example, happy (y1) -- screentime of Instagram (x1) is one pair of well-being outcomes – App-use feature predictors. We have 288 pairs of such outcomes (12 individual outcomes and 4 aggregated outcomes) and predictors (3 features per App for 6 Apps). Figure 4 shows the results for TikTok, WhatsApp, and Instagram. The results for Game Apps, YouTube, and Snapchat are presented in Figure A1.5.

-- For each pair of well-being outcomes – App-use feature predictors, we fit a random intercept model and random slope model. We apply the likelihood ratio test to select the best model following the parsimonious rule.

-- Bonferroni correction is applied to the confidence interval to account for multiple testing. For each well-being outcome and App-use feature, we adjust the level of significance by $0.05/\#$ of tests ($\#=6$) and the confidence interval based on the adjusted level of significance. The confidence interval is adjusted by $Estimate \pm qnorm\left(1 - \frac{0.05}{2*(\#-1)}\right) * Standard_error$, where $\#=6$.

-- The interactive Plotly version of this plot is available in OSF project in path “/code_and_data_for_replication /within_level”.

Figure 1.5. Associations of TikTok, WhatsApp, and Instagram usage with hedonic and eudaimonic well-being outcomes on the same day, the following day (1-day lag), and two days later (2-day lag) in the HappyB study



Note: The Figure shows the unstandardized coefficients and confidence interval between well-being outcomes (Y) and App-specific features (X) of the same day, the following day (1-day lag), and two days later (2-day lag) from mixed linear models, adjusting for well-being outcome of the previous day, age, gender, and school in each model, with time (each day) as level one, and adolescents as level two. Well-being outcomes are

collected from self-reported questionnaires, and App-specific features are derived using text extraction pipeline from the cellphone screenshots of the setting pages uploaded by users, both within the same EMA study.

-- Figure 5 shows the results for TikTok, WhatsApp, and Instagram. The results for Game Apps, YouTube, and Snapchat are presented in Figure A1.6.

-- For each pair of well-being outcomes – App-use feature predictors, we fit a random intercept model and random slope model. We apply the likelihood ratio test to select the best model following the parsimonious rule.

-- **Bonferroni correction is applied to the confidence interval to account for multiple testing.** For each well-being outcome and App-use feature, we adjust the level of significance by $0.05/\#$ of tests ($\#=6$) and the confidence interval based on the adjusted level of significance. The confidence interval is adjusted by $Estimate \pm qnorm\left(1 - \frac{0.05}{2*(\#-1)}\right) * Standard_error$, where $\#=6$.

-- The interactive Plotly version of this plot is available in OSF project in path “/code_and_data_for_replication /within_level”.

Time dynamics of the associations between app usage and well-being

We also applied multilevel models to explain well-being using app-specific features with one-day and two-day lags. The unstandardized correlations for the same-day, 1-day, and 2-day lags are presented for TikTok, WhatsApp, and Instagram (Figure 1.5) and for Game apps, YouTube, and Snapchat (Figure A1.6). We found that most relationships between app-specific features and well-being outcomes remained insignificant over the following one or two days, with the exceptions of Snapchat and TikTok

Specific app usage and adolescent well-being

TikTok. At the between-person level, longer screen time on TikTok correlated with lower levels of contentment ($r=-.18$, $p=.032$), feeling focused ($r=-.18$, $p=.035$), feeling engaged ($r=-.17$, $p=.039$), feeling competent ($r=-.17$, $p=.042$), feeling connected ($r=-.17$, $p=.044$), and eudaimonic activity ($r=-.18$, $p=.029$) across participants, but the associations were not statistically significant after Bonferroni's correction. At the within-person level (Table 1.1), time spent on TikTok was significantly associated with lower eudaimonic well-being ($\beta=-.08$, unstandardized $B=-.04$, $p=.001$) on the same day. In other words, one additional minute of TikTok use corresponded to a .04-point decrease in eudaimonic well-being on a 0-100 scale. In terms of the standardized coefficient, the finding also means that a one SD increases in TikTok use (approximately 39 minutes above an individual's typical daily use) was associated with a 0.08 SD decrease in eudaimonic well-being the same day. Additionally, increased TikTok screen time was linked to lower levels of contentment ($\beta=-.07$, $p=.003$), engagement in daily activities ($\beta=-.06$, $p=.008$), satisfaction ($\beta=-.07$, $p=.007$), and feelings of connection with others ($\beta=-.10$, $p<.001$) on the same day (Figure 1.4), but increased positive emotion the following day ($\beta=.06$, $p=.004$) (Figure 1.5). Moreover, TikTok notifications (one SD=2 notifications above an individual's typical level) were associated with increased negative emotions ($\beta=.06$, $p=.003$) and sadness ($\beta=.07$, $p=.002$) on the same day while TikTok smartphone unlocks were not significantly associated with well-being.

WhatsApp. At the between-person level (Figure 1.3), WhatsApp features showed no statistically significant associations with well-being. At the within-person level (Table 1.1), longer WhatsApp screentime correlated with increased negative emotion ($\beta=.06$, $p=.008$, one SD=30 minutes above an individual's typical level), and this relationship diminished to nearly zero in subsequent day (Figure 1.5). Additionally, WhatsApp smartphone unlocks were related to a greater sense of connection ($\beta=.08$, $p=.008$, one SD=17 activations above an individual's typical level), while the number of WhatsApp notifications was not significantly associated with well-being on the same and following days.

Instagram. At the between-person level, increased Instagram screen activation correlated to higher levels of anger ($r=.21$, $p=.010$) and feeling connected ($r=.21$, $p=.012$), and more Instagram notification was related to enhanced feeling connected ($r=.19$, $p=.010$) and eudaimonic well-being ($r=.15$, $p=.044$). However, these relationships were not statistically significant after Bonferroni's correction. At the within-person level, Instagram notifications was positively related to greater feeling of focused ($\beta=.06$, $p=.006$, one SD=12 notifications above an individual's typical level) on the following day.

YouTube. At the between-person level, more YouTube smartphone unlocks were associated with lower feelings of connection ($r=-.18$, $p=.034$), and more YouTube notifications correlated to higher feelings of satisfaction ($r=.16$, $p=.032$), though neither association was statistically significant after Bonferroni's correction. At the within-person level, on the following day, more YouTube smartphone unlocks were associated with an increased feeling of meaningful ($\beta=.07$, $p=.006$, one SD=2 activations above an individual's typical level); other associations were not significant statistically.

Snapchat. At the between level, more Snapchat smartphone activation correlated to lower happiness ($r=-.15$, $p=.043$) on the same day, but this was not significant statistically after Bonferroni correction. At the within level, no significant association was observed between Snapchat screentime and well-being outcomes on the same day after the Bonferroni correction. However, on the following day, Snapchat notifications (one SD=8 notifications above an individual's typical level) was negatively related to feeling relaxed ($\beta=-.06$, $p=.003$),

feeling content ($\beta = -.07$, $p = .001$), feeling competent in activity ($\beta = -.06$, $p = .002$), feeling satisfied ($\beta = -.07$, $p = .001$), and positive emotion ($\beta = -.06$, $p = .005$).

Game Apps. No relationship was found at the between-person and within-person level after the Bonferroni correction.

2.4 Discussion

The current study employs user-donated screenshots to build individual Appnoms – including different app usage metrics and examines the between- and within-person relationship with hedonic and eudaimonic adolescent well-being collected through EMAs. The discussion about the feasibility, ethical and privacy concerns of the user-donated screenshot and Appnome analysis is presented in another study of our group [99].

According to the Appnome, the most frequently used social media apps were - in order - WhatsApp, Instagram, TikTok, YouTube, Snapchat, and Game Apps (aggregated from various game apps). None of the between-person level relationships with well-being was statistically significant with few exceptions. That aligns with the majority of reviews and meta-reviews that found small and inconsistent effects [18,19,82]. Exceptions included YouTube, where notifications were associated with higher levels of feeling satisfied at the between-person level. Previous studies found a positive correlation between viewing inspiring videos on YouTube and psychological well-being [51], in addition to watching music and infotainment content on YouTube [126]. A review found both negative and positive associations between YouTube use and well-being from both cross-sectional and longitudinal studies [127]. As YouTube is a video browsing app, users primarily passively consume content. Therefore, future studies should aim to unpack the relationship between the content on YouTube and user outcomes. Additionally, at the between-person level, more notifications from Instagram were associated with higher levels of feeling connected and eudaimonic well-being, and more screen activations due to Instagram were also related to a higher level of feeling connected. A previous study found that people who have an Instagram account are less lonely, anxious, and depressed than those who do not have an account [128]. Our findings suggest that the

networking component on Instagram may benefit our study participants by providing online social support and a sense of community [129]. Alternatively, users who have better social connectedness might be more inclined to use Instagram.

At the within-person level, results indicated both positive and negative small effects across platforms. As discussed earlier, different apps are characterized by distinct components, such as profiling, networking, or streaming [58]. This leads users to engage with them in varied ways and contexts to fulfill different needs, which correspond to different psychological processes and well-being outcomes. Below, we discuss the findings for each platform and explore the potential mechanisms associated with each.

At the within-person level, increased screen time on TikTok was associated with decreased levels of eudaimonic well-being, contentment, and feelings of connection with others on the same day, but increased positive emotion the following day. As an app that provides streams of short videos through the “For You” page, users spend more time passively consuming content on TikTok compared to other social media platforms [130]. Therefore, the content and user experience on TikTok play a crucial role in shaping well-being outcomes. According to previous studies, TikTok users report both “enjoyment” and “time distortion” as well as telepresence (immersion in the world of TikTok) [129]. The enjoyment dimension may reflect the benefits of engaging, humorous content and online friendships within TikTok [131]. The time distortion and telepresence dimensions suggest that longer TikTok screen time may serve as an escape from everyday worries, although using TikTok as a coping strategy is suboptimal and can lead to lower eudaimonic well-being on the same day [81,129]. Moreover, since this is a correlational study, it’s important to consider that stressed adolescents might use TikTok as a way to cope with their stress [132].

In addition, increased screen time on WhatsApp was related to elevated levels of negative emotions, which contrasts with previous studies finding that longer time spent on WhatsApp is positively associated with higher psychological well-being [133,134]. As a platform primarily focused on networking and messaging components, WhatsApp offers social support to users but can also increase stress due to information overload or online vigilance,

leading to different well-being outcomes [52]. This discrepancy can be explained by the fact that well-being is affected by individual usage modes and experiences with WhatsApp features. In a previous study, users with more active usage of the read receipt feature reported higher levels of perceived stress [135]. Also, at the within-person level, we found WhatsApp activations were associated with an increased feeling of connection, which is likely related to strengthened bonding social capital [133].

Furthermore, more notifications from Snapchat were not associated with well-being on the same day but were linked to decreased positive emotion, feelings of competence, and satisfaction the following day. Snapchat is a social media platform for multimedia messaging, where people exchange pictures and videos (snaps) that are meant to disappear after being viewed. In previous studies, high Snapchat use was linked to higher delinquent behaviors, school avoidance, and more anxiety and stress [136,137]. One potential explanation for the negative association between Snapchat and well-being is that certain behaviors on the platform increase social comparison. For example, monitoring ex-partners on Snapchat and using the Snap Map feature have been associated with increased jealousy in romantic relationships [137]. Snapchat offers a mix of profiling, networking, and streaming components, compelling researchers to not only link specific user motivations or activities on Snapchat to well-being but also focus on the underlying psychological processes in future studies.

Additionally, we did not find a significant association between well-being and usage of Game apps, measured as an aggregation of various games identified from screenshots. However, it is important to note that "game apps" are an aggregated category, including puzzles, card games, adventure games, and action games, among others. Not all game apps in the app store have social networking aspects. As a result, different types of mobile games can trigger various psychological responses; some alleviate loneliness and boost achievement, while others may lead to addiction among adolescents with low self-esteem due to the game's rewarding nature [138]. Future studies on game apps should ensure a thorough understanding of the features of the games and conduct more focused analyses accordingly.

Overall, our results mainly showed small or non-significant associations. That aligns with the ongoing scientific debate stating that associations might be negligible and speak against the moral panic on social media and adolescent mental health [18,19]. Besides the rise of social media, other factors like a broader definition of mental health and greater awareness of mental health issues could also explain the decline in adolescent mental health [139]. Previous research has also shown approaches through which social media can benefit youth mental health [140]. In the current study, though shifting from aggregated social media use metrics towards within-person variations offers deeper insights into the underlying mechanisms of social media's impact on adolescent well-being [54,101,141], the small size of the effects showed that these features are still insufficient to fully capture the social media behaviors or activities that risk or promote well-being. One practical implication is that intervening on these features might not have a great effect on well-being.

The fact that many associations were not statistically significant after applying a more restrictive threshold for significance highlights the importance of adjusting for multiple hypothesis testing when studying app-specific data. Correcting for multiple testing is essential to prevent p-hacking and inaccurate results due to chance when exploring a large amount of data [142]. It bears noting, however, that Bonferroni correction is conservative by design, and in the context of an exploratory study, this stringency carries a real cost. The correction substantially increases the risk of false negatives. Specifically, associations with genuine empirical relevance may not survive the corrected threshold not because they are absent, but because the study is underpowered to confirm them at this level of strictness. This issue is particularly consequential given that much research in the social media and well-being domain, including the present study, is better characterized as hypothesis-generating than hypothesis-testing. In hypothesis-generating work, the goal is to identify candidate associations worthy of further investigation, not to achieve definitive confirmation. Therefore, in hypothesis-generating work, a less conservative significance threshold is more appropriate for flagging preliminary signals. We therefore suggest that associations in the present study which show consistent directional patterns across multiple outcomes or platforms, but do not survive strict

multiple comparison correction, should not be dismissed as pure nulls. Future studies in this domain would benefit from explicit attention to two design decisions often made implicitly: whether the analysis sits in a hypothesis-generating or hypothesis-testing paradigm, and whether the chosen correction procedure is calibrated to the study scope and power estimation.

Finally, we examined the temporality of within-person associations between app-specific usage features and well-being. In cases where within-person associations were almost significant on the same day, most relationships diminished to near zero the following day. Overall, these results raise an important point concerning the temporality of the effects of social media [85,106] and underline that these relationships can change over time. We can speculate that these temporal effects can be due to several reasons. On one hand, the impact of social media might be cumulative rather than instant. Repeated exposure to certain types of content or interactions could gradually influence well-being, becoming more apparent a day later as these effects build up. On the other hand, the feedback from social media, such as comments and likes, might not be immediately seen but reviewed the next day, thus influencing well-being once the user has had time to absorb and react to this feedback. These temporal changes should be considered even when policymakers need to act.

Limitations

The current study has three main limitations. First, the study experienced a relatively high attrition rate. Only 74.1% (277 out of 374) of the participants who completed the baseline survey finished at least one assessment and submitted screenshots along with the assessment among the total 14-day assessments. This attrition rate falls within the slightly lower range of those reported in existing EMAs [143,144]. Our attrition analysis showed that the drop-off group and the group with good adherence were generally comparable, except that female participants were more likely to remain in the sample (Table A1.3). However, despite the attrition analysis results, the high level of attrition may introduce the issue of missing not at random, particularly regarding social media use or other socio-demographic characteristics

that we did not measure. It is also important to note that the attrition analysis was conducted on demographic and baseline characteristics, but not on the study outcomes themselves as participants who dropped out are no longer providing outcome data, their well-being trajectories during the study period remain unobserved. This means that systematic differences in well-being or social media use between those who remained and those who attrited cannot be ruled out, and the analytic sample may over-represent participants with particular patterns of engagement or well-being that differ from those who left the study early.

Second, participants who consistently donated screenshots throughout the study may differ systematically from those who did not, for example in their engagement with social media, comfort with data sharing, or general mindset about social media use, introducing a further layer of selection bias that may limit the representativeness of the analytic sample. Additionally, the OCR-based extraction of app usage data from screenshots may introduce non-random measurement error, as recognition accuracy can vary systematically across device types, operating system versions, and app interface designs, potentially creating uneven measurement quality across participants.

Third, the observed app-specific usage time from user-donated screenshots is subject to a missing data problem. Only the top three apps' usage features are observed in the screenshots; the fourth app and beyond are not. In the current analysis, we assume that the usage time of an unobserved app is zero. However, this assumption may be incorrect, as the app could have been used for some time. Beyond this coverage limitation, even for the three observed apps, the extracted metrics are aggregates that treat qualitatively distinct experiences as equivalent. Passive browsing and active interaction, or harmful content consumption and peer support seeking, contribute equally to a screen time total, obscuring the contextual and content-dependent variation that the broader dissertation argues is central to understanding social media's effects on well-being.

Fourth, the current study can be improved by considering the reciprocal relationship between social media use and well-being through advanced modeling approaches such as cross-lagged models and structural equation models. Additionally, the linear mixed-effects

models employed here assume linear, additive associations and fixed lag structures, which may not fully capture nonlinear or dynamic processes inherent to social media use, such as threshold effects, dose-response relationships, or feedback loops between use and well-being over time. Given the scope of the current study, we did not conduct these additional analyses, but we plan to explore these approaches in future research.

Finally, the sample consists of Swiss adolescents recruited from four high schools in Canton Ticino, and is not representative of the broader adolescent population. Findings may not generalize to adolescents in other national, cultural, or socioeconomic contexts, and future research should examine whether app-specific associations with well-being hold across more diverse populations.

2.5 Conclusion

The current study offers a novel approach to study the relationships between Appnomes and hedonic and eudaimonic adolescent well-being. Our results highlight the importance of studying both between- and within-person variations, as well as the temporal dynamics in the association between social media use and well-being.

Supplementary materials

Figures

Figure A1.1. Study Design of Chapter 2

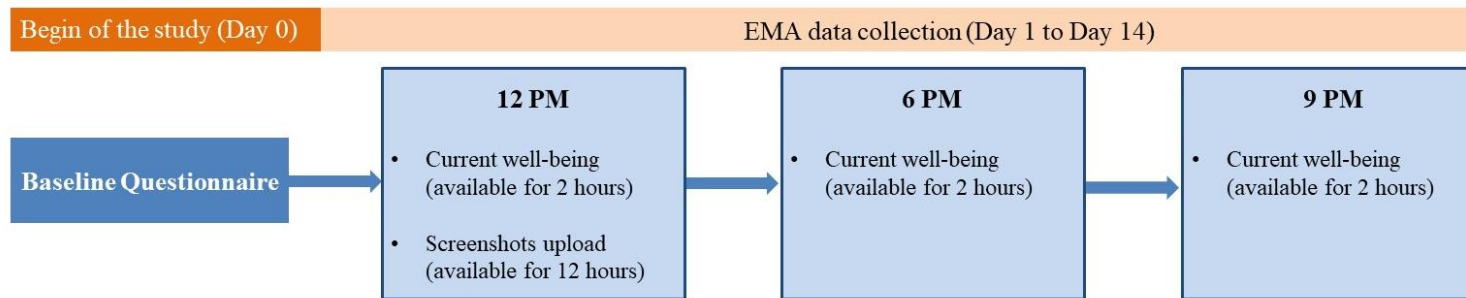


Figure A1.2. Examples of screenshots on number of screentime (A), number of screen unlock/activation (B), and number of notifications (C)

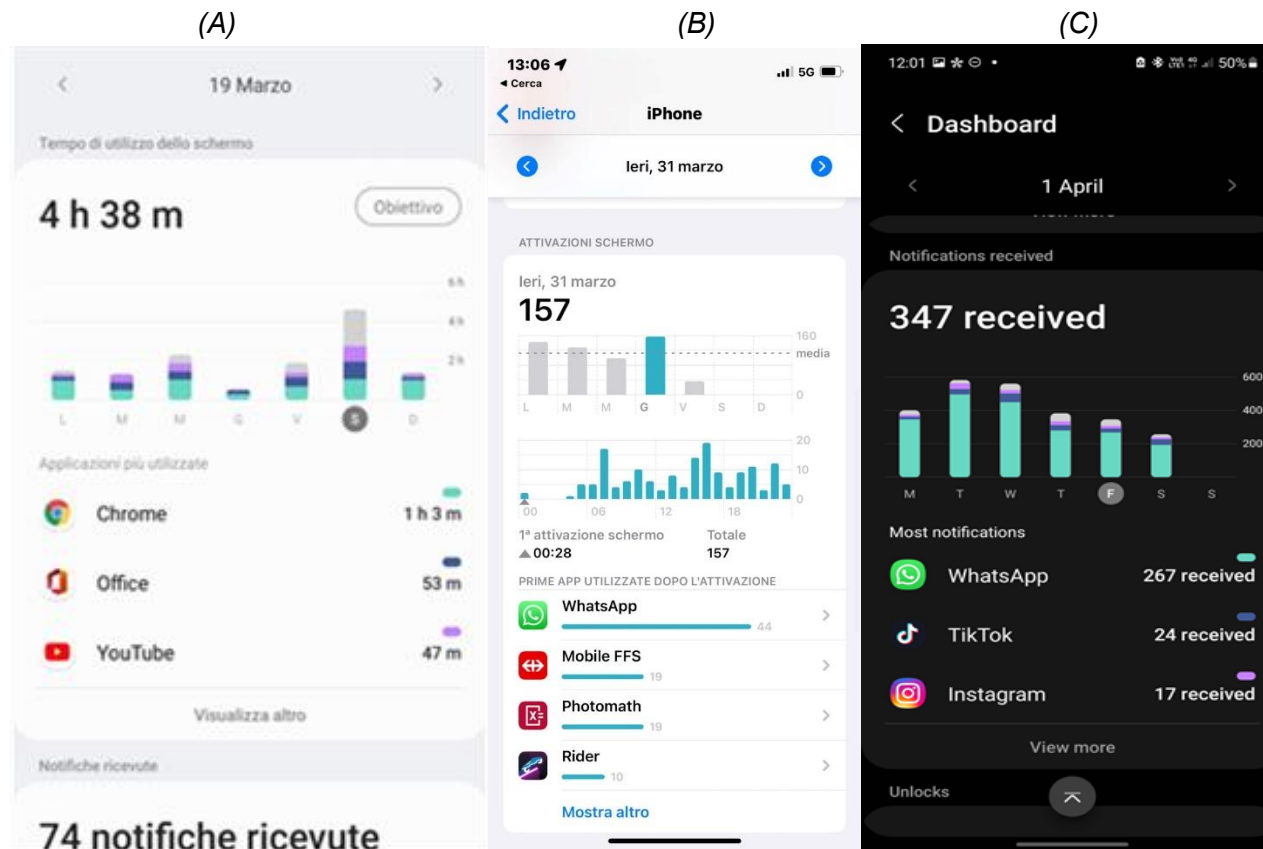
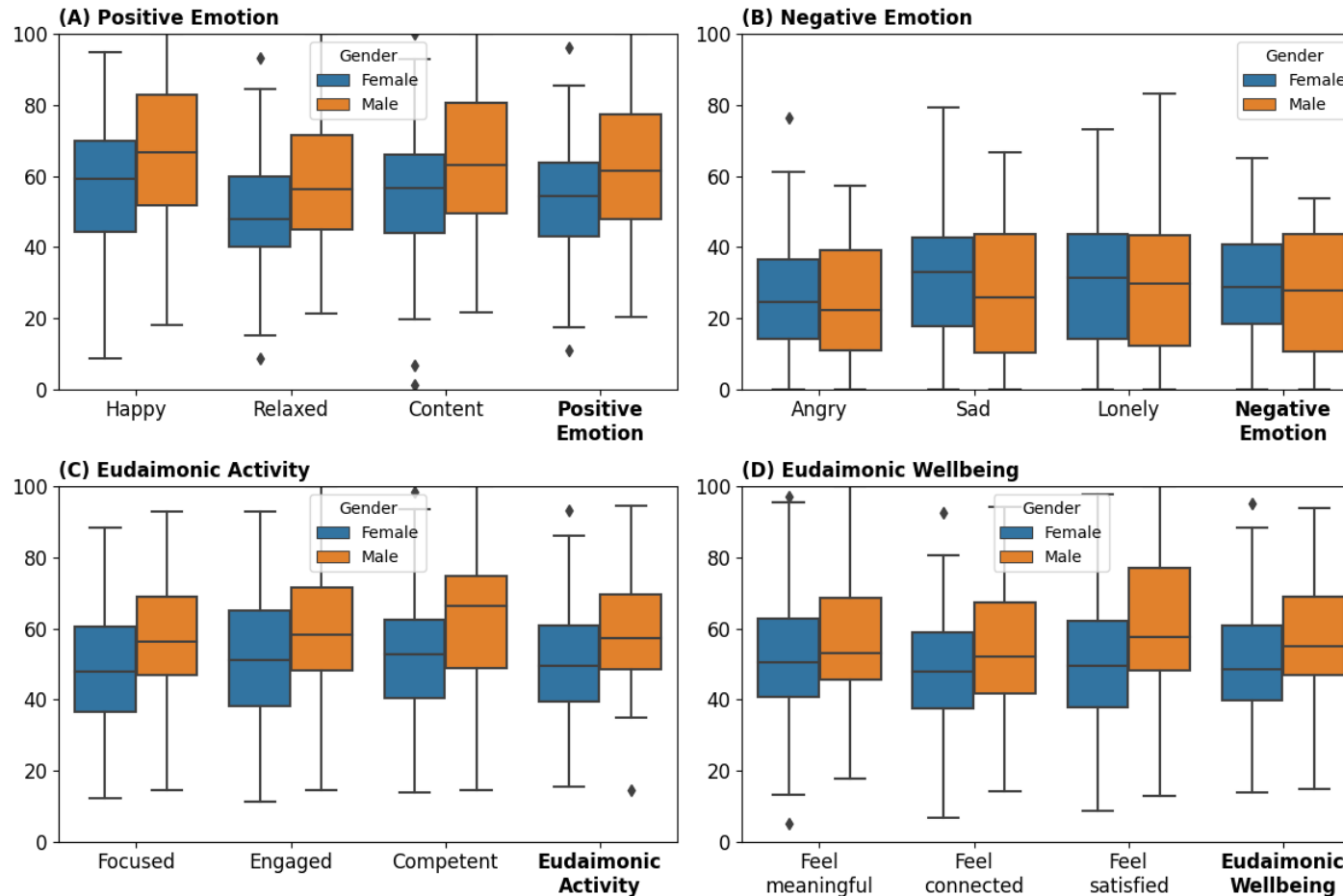


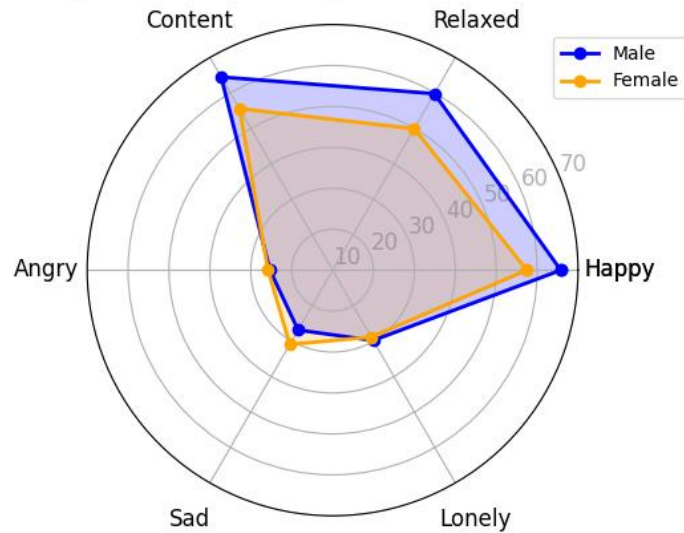
Figure A1.3. The distribution of person-level average positive emotion (A), negative emotion (B), eudaimonic activity (C), and eudaimonic well-being (D) by gender in the HappyB study



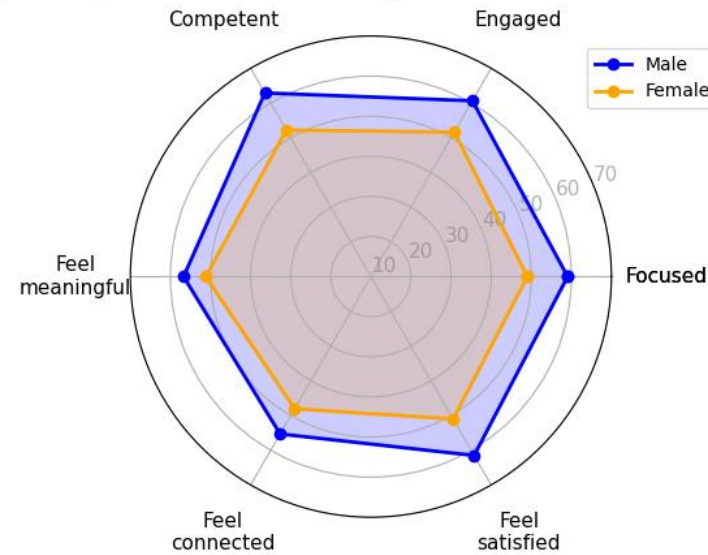
Note: The figure presents the distribution of the person-level well-being outcomes. Participants completed a daily self-reported survey on hedonic and eudaimonic well-being throughout the 14-day study period. Positive emotion is measured by the level of Happiness, Relaxation, and Contentment, while Negative emotion is assessed based on Anger, Sadness, and Loneliness. Eudaimonic well-being is measured by the level of feeling meaningful, connected, and satisfied. Eudaimonic activity is measured by the level of feeling engaged, competent, and focuses

Figure A1.4. The radar plot of the hedonic (A) and eudaimonic well-being (B) outcomes by gender

(A) Average Hedonic Well-being Score

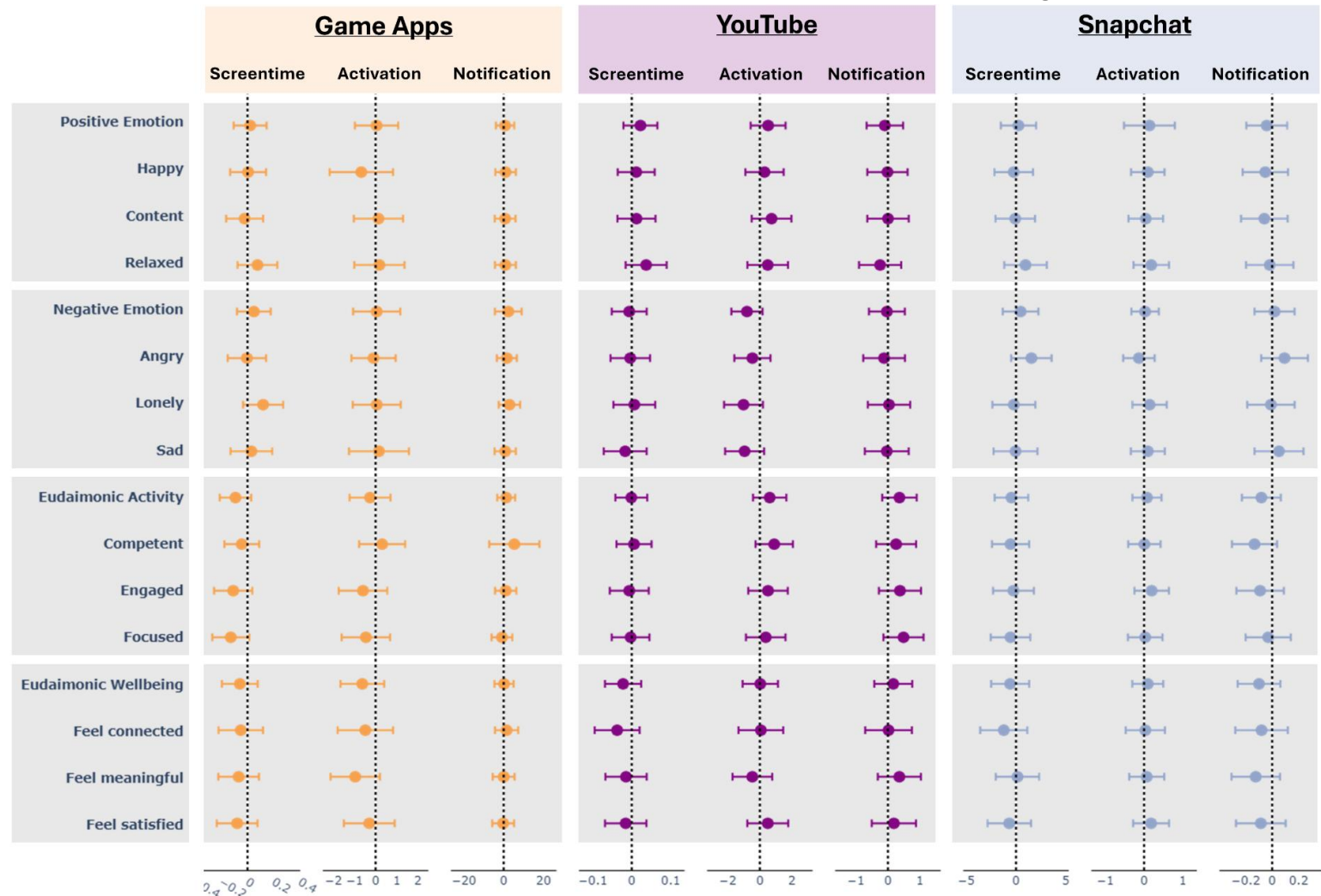


(B) Average Eudaimonic Well-being Score



Note: The figure presents the distribution of the person-level well-being outcomes. Participants completed a daily self-reported survey on hedonic and eudaimonic well-being throughout the 14-day study period. Hedonic well-being is measured by the level of Happiness, Relaxation, Contentment, Anger, Sadness, and Loneliness. Eudaimonic well-being is measured by the level of eudaimonic well-being, including feeling meaningful, feeling connected, and feeling satisfied, and the level of eudaimonic activity, including the level of feeling engaged, competent, and focused.

Figure A1.5. Forest plot of the mixed linear models results on the associations between Games Apps, YouTube, and Snapchat screentime, number of activations, and number of notifications and hedonic/eudaimonic well-being



Estimated associations between app usage and well-being outcomes

Note:

-- S-Figure 1 shows the unstandardized coefficients and confidence interval between well-being outcomes (Y) and App-specific features (X) of the same day from mixed linear models, adjusting for well-being outcome of the previous day, age, gender, and school in each model, with time (each day) as level one, and adolescents as level two. Well-being outcomes are collected from self-reported questionnaires, and App-specific features are derived using text extraction pipeline from the cellphone screenshots of the setting pages uploaded by users, both within the same EMA study.

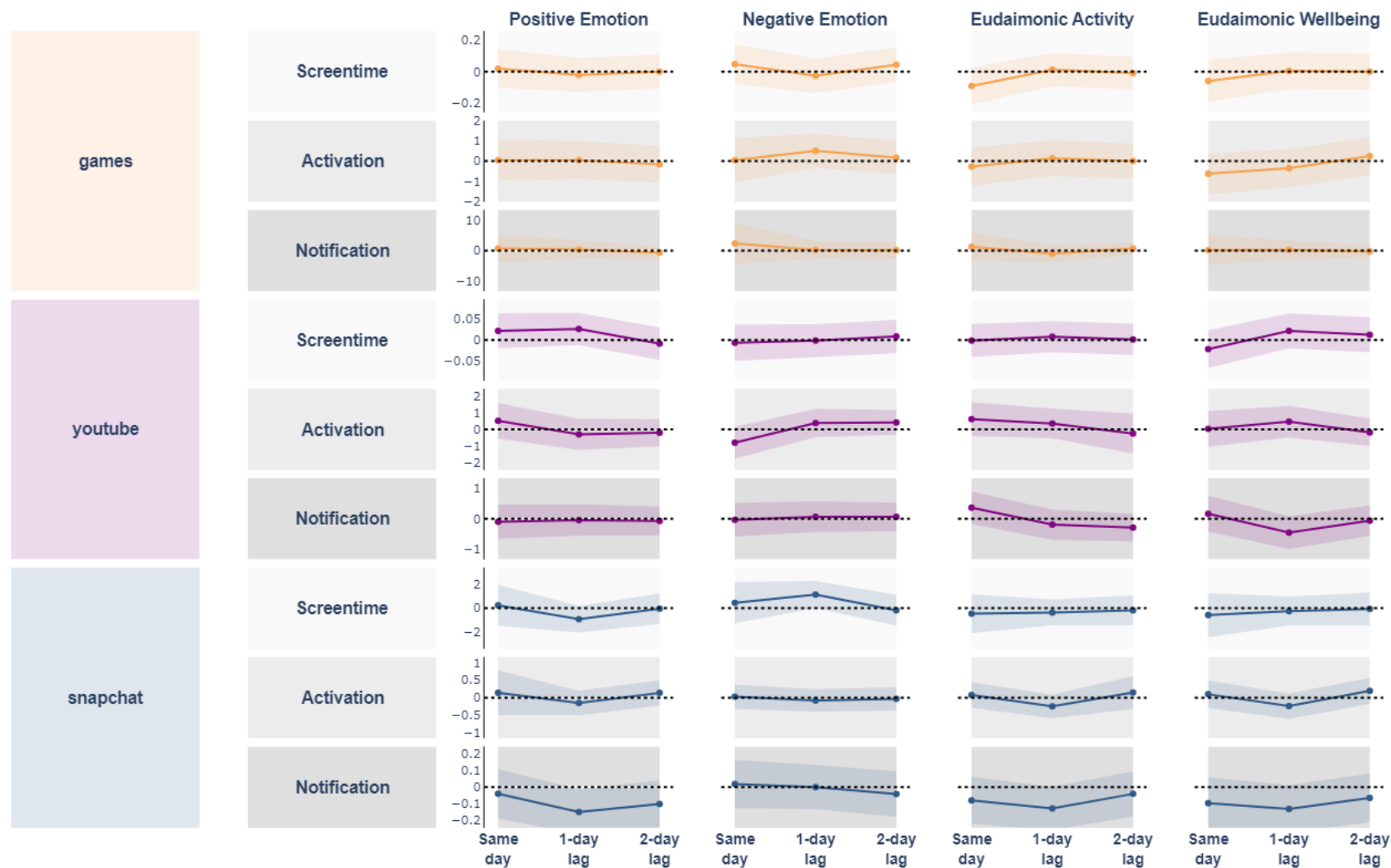
-- For example, happy (y1) -- screentime of Instagram (x1) is one pair of well-being outcomes – App-use feature predictors. We have 288 pairs of such outcomes (12 individual outcomes and 4 aggregated outcomes) and predictors (3 features per App for 6 Apps).

-- For each pair of well-being outcomes – App-use feature predictors, we fit a random intercept model and random slope model. We apply the likelihood ratio test to select the best model following the parsimonious rule.

-- **Bonferroni correction is applied to the confidence interval to account for multiple testing.** For each well-being outcome and App-use feature, we adjust the level of significance by $0.05/\#$ of tests ($\#=6$) and the confidence interval based on the adjusted level of significance. The confidence interval is adjusted by $Estimate \pm qnorm\left(1 - \frac{0.05}{2*(\#-1)}\right) * Standard_error$, where $\#=6$.

-- The interactive Plotly version of this plot is available in OSF project in path “/code_and_data_for_replication /within_level”.

Figure A1.6. Associations of Games Apps, YouTube, and Snapchat usage with hedonic and eudaimonic well-being outcomes on the same day, the following day (1-day lag), and two days later (2-day lag), results from mixed linear models



Note:

-- S-Figure 2 shows the unstandardized coefficients and confidence interval between well-being outcomes (Y) and App-specific features (X) of the same day, the following day (1-day lag), and two days later (2-day lag) from mixed linear models, adjusting for well-being outcome of the previous day, age, gender, and school in each model, with time (each day) as level one, and adolescents as level two. Well-being outcomes are collected from self-reported questionnaires, and App-specific features are derived using text extraction pipeline from the cellphone screenshots of the setting pages uploaded by users, both within the same EMA study.

-- For each pair of well-being outcomes – App-use feature predictors, we fit a random intercept model and random slope model. We apply the likelihood ratio test to select the best model following the parsimonious rule.

-- **Bonferroni correction is applied to the confidence interval to account for multiple testing.** For each well-being outcome and App-use feature, we adjust the level of significance by $0.05/\#$ of tests ($\#=6$) and the confidence interval based on the adjusted level of significance. The confidence interval is adjusted by $Estimate \pm qnorm\left(1 - \frac{0.05}{2*(\#-1)}\right) * Standard_error$, where $\#=6$.

-- The interactive Plotly version of this plot is available in OSF project in path “/code_and_data_for_replication /within_level”.

Tables

Table A1.1. The accuracy rate of text extraction in HappyB screenshot analysis

	Screentime (IOS)	Screentime (Android)	Activation (IOS)	Activation (Android)	Notification (IOS)	Notification (Android)
Total screentime/count	0.91	0.93	0.91	0.94	0.95	1.00
APP #1 name	0.98	0.91	0.97	0.99	0.91	0.97
APP #1 screentime/count	0.94	0.87	0.94	0.99	0.89	0.94
APP #2 name	0.90	0.89	0.96	1.00	0.91	0.94
APP #2 screentime/count	0.91	0.88	0.93	1.00	0.91	0.94
APP #3 name	0.89	0.92	0.95	1.00	0.88	0.94
APP #3 screentime/count	0.86	0.87	0.95	0.99	0.89	0.93

Note: The Table shows the measurement-level accuracy rate of text extraction in the HappyB screenshot analysis. The measurement-level accuracy rate is determined by the proportion of screenshots for which the extracted text matches the ground truth exactly, in terms of measurements such as total screen time, relative to the total number of screenshots within that category. It's important to recognize that the high accuracy rate observed in the Android activation category can be attributed to many screenshots in our dataset originating from an older Android version.

Table A1.2. Descriptive analysis for well-being outcomes and app-specific usage features by gender and age

	Gender		P value	Age				P value
	Male (N=56)	Female (N=148)		15 (N=97)	16 (N=72)	17 (N=25)	18 (N=8)	
Well-being outcomes								
Positive Emotion	63.36	54.32	0.007	57.53	56.52	54.63	47.34	0.551
Negative Emotion	27.34	28.53	0.714	23.89	32.15	30.22	40.17	0.009
Eudaimonic Activity	60.86	50.95	0.002	54.38	53.09	51.96	44.83	0.583
Eudaimonic Wellbeing	57.85	50.09	0.019	52.12	53.68	47.96	44.08	0.435
Average Screenshot by app								
Games	3.44	4.33	0.684	4.08	5.24	1.86	0.85	0.636
Instagram	27.85	37.89	0.168	33.57	34.83	36.56	63.56	0.293
Snapchat	0.05	0.19	0.508	0.04	0.24	0.50	0.00	0.459
TikTok	24.93	78.19	0.000	66.51	64.09	70.69	59.13	0.976
WhatsApp	25.11	36.99	0.067	29.59	37.89	50.03	14.46	0.049
YouTube	44.27	17.90	0.001	30.29	20.67	14.47	5.40	0.284
Average number of activations by app								
Games	0.49	0.25	0.267	0.28	0.38	0.18	0.12	0.838
Instagram	11.39	15.02	0.338	15.61	12.66	11.70	23.98	0.361
Snapchat	0.85	1.40	0.478	1.37	0.90	2.01	1.56	0.684
TikTok	1.18	3.89	0.027	4.28	3.11	1.67	2.22	0.247
WhatsApp	23.18	32.90	0.082	30.88	30.29	35.78	26.83	0.820
YouTube	1.16	0.44	0.061	0.69	0.58	0.17	0.41	0.690
Average number of notifications by app								
Games	0.10	0.09	0.962	0.12	0.04	0.19	0.00	0.601
Instagram	5.84	15.93	0.003	12.80	14.24	15.69	9.72	0.856
Snapchat	7.62	12.17	0.247	10.84	11.38	12.49	7.56	0.961
TikTok	0.33	0.64	0.490	0.52	0.45	1.21	0.17	0.625
WhatsApp	106.83	114.07	0.633	99.93	123.80	135.47	97.27	0.194
YouTube	2.15	0.72	0.229	1.75	0.53	0.13	0.00	0.594

Note: P-values for outcomes by gender are derived from the t-test, and p-values for the outcomes by age are derived from ANOVA.

Table A1.3. Attrition analysis of the current study

		Finish < 2 assessments (N=169)		Finish 2-10 assessments (N=151)		Finish >=10 assessments (N=54)		Test Statistics	P value
		Mean	SD	Mean	SD	Mean	SD		
	Age	15.77	0.83	15.72	0.84	15.60	0.79	0.66	0.519
	Female (%)	53.85		74.35		65.08		7.06	0.029
Screentime	Games	6.45	36.51	5.31	21.29	4.56	20.12	0.33	0.717
	Instagram	51.06	56.74	35.85	48.07	34.76	42.78	1.94	0.144
	Snapchat	0.70	3.94	0.72	7.11	0.24	3.39	1.34	0.263
	TikTok	74.24	120.55	66.14	81.55	55.33	72.02	4.28	0.014
	WhatsApp	35.52	43.58	36.83	50.94	37.67	49.69	0.07	0.928
	YouTube	27.70	42.34	25.85	55.84	26.67	52.10	0.06	0.941
Number of activation	Games	0.45	1.89	0.51	2.65	0.52	2.35	0.01	0.990
	Instagram	19.52	21.71	14.37	21.12	14.33	20.07	1.00	0.369
	Snapchat	0.30	1.71	1.66	7.45	1.07	4.29	2.16	0.116
	TikTok	4.52	11.42	4.45	9.33	5.06	9.47	0.83	0.437
	WhatsApp	30.18	47.35	32.93	34.67	37.29	34.88	3.29	0.038
	YouTube	0.00	0.00	0.77	3.30	0.34	2.21	4.99	0.007
Number of notification	Games	0.26	1.43	0.25	2.22	0.14	1.35	0.88	0.416
	Instagram	10.23	18.87	14.89	30.55	12.85	23.97	1.64	0.194
	Snapchat	5.67	13.54	12.73	53.63	7.89	22.37	3.14	0.043
	TikTok	0.00	0.00	0.96	5.58	1.32	6.88	1.46	0.232
	WhatsApp	95.08	90.25	122.14	124.44	134.34	141.97	3.25	0.039
	YouTube	0.26	1.13	1.91	10.94	2.23	11.56	0.69	0.501

Note: The Table presents the average age, proportion of females, and the average App-specific features among participants who finished less than 2 assessments, who finished 2-10 assessments, and more than 10 assessments. Analysis of variance (ANOVA) is applied to compare group means for age and App-specific features, and the Chi-2 test is applied to compare the proportion of females among the three groups.

Chapter 3: Higher goal awareness improves social media user experience and well-being

Authors

Yuning Liu¹, Amit Goldenberg²

Affiliations

¹ Harvard Kenneth C. Griffin Graduate School of Arts and Sciences, Harvard T.H. Chan School of Public Health

² Harvard Business School

Abstract

A key challenge in social media use and well-being research is that social media behaviors are highly context- and content-dependent. This study therefore examines an underlying psychological process by conceptualizing social media use as momentary goal pursuit. Specifically, it investigates how goal awareness, defined as awareness of one's purpose for using social media, and goal strength, the degree of motivation to pursue that goal, shape momentary user experiences and well-being. Using a 7-day ecological momentary assessment (EMA) design with 266 U.S. adults (six surveys per day; 71.3% response rate), the study integrates self-reported goals for social media use and well-being outcomes, screentime screenshots, and user-donated TikTok behavioral data. Within individuals, higher goal awareness and goal strength were associated with more positive experiences and better well-being, respectively, particularly when social media use served goals such as coping, entertainment, passing time, or habitual use. TikTok use showed little behavioral variation by goal awareness and goal strength, except that users with higher goal awareness skipped videos faster. From pre- to post-EMA, nearly half of participants reported increases in perceived agency in their social media use and more positive

experiences. Findings highlight goal awareness as a key process shaping momentary social media use experiences, while also underscore the need to better capture behavioral signals of intentional social media use.

3.1 Introduction

Social media has become an integral part of daily life. In 2025, 84% of U.S. adults reported using YouTube and 71% reported using Facebook [145]. Globally, social media platforms engage billions of users, with especially high uptake among adolescents [145–147]. Notably, the global expansion of social media use has coincided with population-level declines in mental well-being, yet the causal nature of this relationship remains unresolved [17,148]. The global number of disability-adjusted life-years (DALYs) attributable to mental disorders increased from 80.8 million to 125.3 million from 1990 to 2019 [10], with an estimated 5.7% of adults globally suffer from depression, and suicide remains the third leading cause of death among those aged 15–29 [11]. In light of these trends, a growing body of research has examined whether social media use is associated with mental health.

When examining the association between time spent on social media and well-being, systematic reviews and meta-analyses have reported either small, inconsistent, or heterogeneous effects [18,19,27]. This lack of a clear effect, has led researchers to try and establish clarity on the association by zooming in on specific types of social media use to identify more fine-grained behaviors that may be particularly harmful or beneficial. In line with the goal of associating well-being with more specific social media behavior, prior studies have examined different aspects of use such as type of use behavior (e.g., active versus passive use, or specific behaviors such as browsing or posting) [36–38], content exposure (e.g., misinformation, mental health–related content, online social support, hate speech) [44,45], and the physical or social context of use [54]. We refer to these collectively as type-specific approaches, which are studies that focus on particular dimensions of social media use, whether defined by time, activity, content, context, or

platform. Type-specific approach is well suited for identifying dimension-specific risks and benefits of social media use. However, individuals typically engage in multiple forms of social media use, shifting across platforms, activities, and contexts constantly. As a result, type-specific approaches in the study of social media use and well-being may be limited in integrating effects across different use types to interpret individual-level heterogeneity, and informing feasible interventions for users, designers, or policymakers that must operate across diverse and evolving types of use. Yet comparatively little research has focused on identifying insights that generalize across different types of social media use. Therefore, in this study, instead of subdividing types of social media use, we adopt a type-independent approach by focusing on the psychological processes that operate across social media use.

In this study, we conceptualized social media use as a process of momentary goal pursuit. Inspired by theories of goal pursuit [149–151], we focus on two factors that shape the effectiveness of goal pursuit in social media use: goal strength, defined as the momentary intensity of motivation, and goal awareness, defined as users' conscious awareness of their goals during use. While both goal strength and goal awareness have been shown to facilitate effective goal pursuit in other domains [152–154], these constructs, particularly goal awareness, remain underexamined in social media research. This approach holds promise because in attention-optimized social media environments [155,156], users' momentary goals can easily drift during use, making goal strength and goal awareness central processes for understanding experience across types of use. In light with these ideas, the present study examines how momentary goal strength and goal awareness shape social media experiences and well-being.

Social media use and well-being

Despite the common lay perception that social media use is inherently harmful, empirical research examining the relationship between social media use and well-being has yielded mixed findings [18,19,27,82]. The mixed findings likely reflect variation in study populations, designs,

disciplinary approaches, platforms, and outcomes examined [26,83]. More fundamentally, studies examining different types of social media use face the challenge that individuals rarely engage in only one type of use. Instead, they routinely move across platforms, activities, and contexts within short periods of time [74,157,158].

One emerging line of research emphasizes users' subjective perceptions of their social media experiences, independent of specific types of social media use [53,159,160]. For example, research on social media mindsets shows that perceived agency over use (in control vs. out of control) and perceived valence of its effects (enhancing vs. harmful) are associated with well-being outcomes [53,160]. Notably, the associations between mindset and psychological distress are not mediated by specific types of use [53], suggesting that social media mindsets represent psychological states that may generalize across different types of social media use. Building on social media mindset research, which captures users' global, trait-level evaluations of agency and valence, the present study extends this perspective to momentary fluctuations in psychological states and shifts focus from reflective assessments of use to goal-pursuit processes that precede and accompany social media engagement.

Social media use as goal pursuit process

We conceptualize social media use as a sequence of momentary goal-pursuit episodes. This perspective builds on two lines of research on the motives underlying social media use and habitual social media use. Empirical evidence suggests that social media use is goal driven. Prior research has identified a wide range of motives for social media use, including self-expression, networking, social interaction [161–163], entertainment [164], stress coping [165], and filling idle moments in everyday routines [166]. Alongside this line of work, habitual social media use has been widely observed, such as repeated engagement with specific platforms [167] or routine activities such as checking or posting [168,169]. Although habits are often characterized as automatic, they are not independent of goals. Rather, habits emerge through repeated goal

pursuit in stable contexts, where cue–routine–reward cycles become reinforced over time [170,171]. Thus, seemingly habitual social media behaviors can be understood as learned forms of goal pursuit that are initiated by contextual cues. Therefore, both perspectives indicate that social media use reflects goal pursuit. Moreover, motives for social media use vary both between individuals and within individuals over time, as users shift goals in response to situational demands [157,158,172]. Habitual use is likewise dynamic, comprising nested sub-habits within broader routines [57]. Accordingly, social media goal pursuit is best understood not as a single, stable behavior, but as a sequence of momentary goal-pursuit episodes.

Goal strength

When considering goal pursuit, one central dimension is goal strength, referring to the intensity or strength of a goal [149,173–175]. Some initial social media research documents variation in the intensity of goals to engage with social media. For example, intentions to socialize online vary across individuals, and stronger intentions predict greater subsequent engagement [176,177]. Longitudinal studies also show within-person daily fluctuations in goal-related processes such as need satisfaction and frustration in social media use [178], and further found goal conflict moderated the relationship between online vigilance and perceived stress [179]. In addition, empirical research shows that stronger goal strength is associated with more effective goal achievement [152]. Across domains such as education, health behavior, and emotion regulation, research shows that stronger goal strength predicts greater sustained effort and more positive behavioral and well-being outcomes [153,180,181]. Extending these insights to social media, we expect that stronger momentary goals promote more deliberate, goal-directed engagement (e.g., seeking humorous content to unwind or searching for a friend’s post to reconnect), whereas weaker goals may increase reliance on default affordances such as algorithmically curated feeds, resulting in mindless or maladaptive use. Accordingly, we

hypothesize that stronger goals during social media use facilitate goal achievement and, in turn, more positive user experiences and well-being outcomes.

Goal awareness

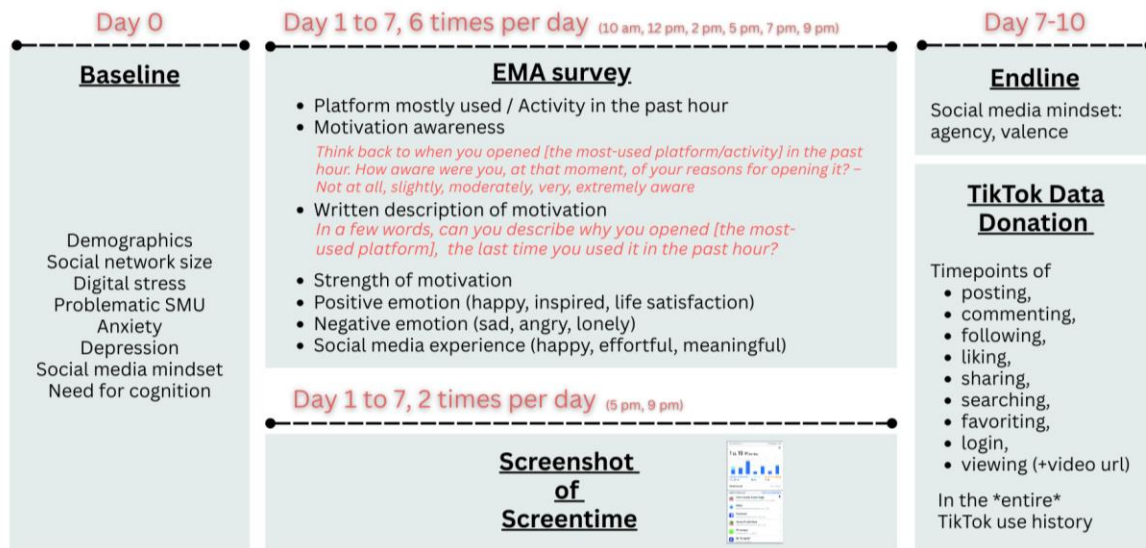
Beyond goal strength, another important but less studied dimension of goal pursuit is goal awareness. In the context of social media use, goal awareness refers to the extent to which users are consciously aware of their goals during momentary engagement. Social media goal pursuit can occur with reduced conscious awareness, particularly because of the habitual nature of social media use as described above. For example, users report experiences such as doomscrolling, time distortion, “rabbit-hole” engagement, as well as late-night use despite fatigue, in which they initiate social media with little awareness and only realize time has passed after prolonged sessions [40,41,182–184]. Although conscious awareness is not always necessary for goal achievement, it may be particularly important for goal attainment when goal pursuit is effortful or challenging [150,185–187]. Specifically, conscious awareness supports goal pursuit by directing attention toward goal-relevant information and mobilizing cognitive resources [150]. Experimental research shows that increasing the salience and conscious accessibility of goals enhances persistence, performance, and emotional experience, particularly under conditions of goal conflict [188]. Consistent with this view, ecological momentary assessment studies show that greater momentary awareness of the source of negative emotions is associated with more frequent emotion regulation attempts, greater use of effective strategies, and higher perceived regulation success [154,189]. Extending these insights to social media use, we argue that a higher level of awareness of one’s goals enables more effective and adaptive engagement with social media, thereby supporting more positive user experiences and well-being.

The current study

To test this account, we conducted an EMA study that repeatedly measured motivation awareness, motivation strength, motivation type, and well-being in daily life. These self-reports were linked to objective TikTok use logs and screenshots of device-level screen time, allowing us to examine psychological processes alongside observed behavior. We focused on TikTok for objective behavioral data collection for two reasons. First, short-form video has become the dominant mode of social media consumption, with TikTok reaching over one billion active users globally and representing one of the fastest-growing platforms among young adults [190]. Its algorithmically curated feed, designed to maximize continuous engagement, makes it a particularly relevant context for examining goal pursuit processes and their behavioral signatures. Second, TikTok's data donation package provides structured and accessible forms of objective behavioral data, including timestamped records of individual video interactions, making it well-suited for linking behavioral patterns to EMA-reported psychological states. The current study aims to test the following hypotheses:

- H1: At both the between-person and within-person levels, higher goal awareness and goal strength in social media use are associated with more positive user experiences and better well-being outcomes, respectively.
- H2: The associations in H1 holds regardless of the specific type of the goal of social media use.
- H3 (explanatory): Higher goal awareness and goal strength are associated with activities and experience on TikTok that align more closely with an individual's goals.

Figure 2.1. The study design for the goal awareness and social media experience study



3.2 Method

3.2.1 Sample recruitment

We conducted a 7-day EMA study to test the proposed hypotheses. Participants were recruited via Prolific, an online research recruitment platform. Eligibility criteria required participants to be English-speaking adults residing in the United States who had used TikTok at least once in the past month. Data collection was conducted in July 2025. A total of 266 participants were recruited (Mean age = 37.9 years; 61.28% female). All participants provided informed consent prior to participation.

3.2.2 Study procedure

The study design was presented in Figure 1. After recruitment on Prolific, participants were directed to complete a baseline survey administered via Qualtrics. Following the baseline survey, participants were guided to download a smartphone-based EMA app, Expiwell, on their personal devices. Participants then completed a 7-day EMA protocol. Participants received six EMA prompts per day for seven consecutive days (Days 1–7), delivered at fixed times: 10:00 a.m.,

12:00 p.m., 2:00 p.m., 5:00 p.m., 7:00 p.m., and 9:00 p.m. Each prompt asked participants to reflect on their experiences during the preceding hour. To obtain objective indicators of screen use, participants were instructed to submit screenshots of their smartphone screen-time statistics twice daily, at the 5:00 p.m. and 9:00 p.m. surveys, across the 7-day period. Screenshots were taken directly from the device's system settings and included total screen time and app-specific usage, allowing for objective assessment of social media use duration.

At the end of the EMA period, participants completed an endline survey and were invited to share their TikTok data through a secure data donation platform. Specifically, participants were guided to download their TikTok data donation package from either the TikTok mobile application or web account interface. TikTok data donation packages are personal data archives containing a user's complete activity history on the platform, including video viewing records, interaction behaviors (e.g., likes, shares, comments), and usage timestamps, which platforms are legally required to make available to users under data privacy regulations such as the EU General Data Protection Regulation (GDPR) and comparable U.S. frameworks. Participants then uploaded the data package (ZIP file) to the secure Next-Eyra data donation platform. In the Eyra platform, personally identifiable information in the TikTok data package is removed prior to researcher access. Participants were also able to select which categories of TikTok data and what data they wished to share with researchers. Only de-identified data explicitly consented to by participants were made available to the research team. Participants were compensated via Prolific in accordance with platform standards and study completion.

3.2.3 Study measurements

Independent variables

At each EMA prompt, participants first reported the platform they used most in the past hour. If they indicated that they had not used any social media platform, they were asked to select the primary non-social media activity they engaged in during that hour (e.g., studying, shopping,

or household tasks). When participants reported using a social media platform in the past hour, they were asked to report their momentary goal pursuit during that period, including goal awareness, a written description of their goal, and goal strength. Goal awareness was assessed with the question, “Think back to when you opened [the most-used platform/activity] in the past hour. How aware were you, at that moment, of your reasons for opening it?”, using a 5-point Likert scale ranging from not at all aware to extremely aware. Participants then provided a brief open-ended description of their motivation (“In a few words, can you describe why you opened [the most-used platform], the last time you used it in the past hour?”). Goal strength was measured with the item, “In the past hour, when you used social media, how motivated were you to achieve the reason you described in the previous question?”, using the same 5-point Likert scale. When participants reported no social media use in the past hour, the same set of questions assessing goal awareness, description, and strength was administered with reference to the specific non-social media activity they reported engaging in most during that period.

Dependent variables

At each EMA prompt, we assess the two sets of dependent variables: subjective well-being and social media use experience. Momentary subjective well-being was assessed at each EMA prompt using self-reported emotional states. Participants rated how strongly they were feeling several emotions at the moment of the survey on a scale from 0 (not at all) to 100 (extremely). Positive emotions included happiness, inspiration, and life satisfaction, while negative emotions included distress, anger, and loneliness. Items were analyzed as indicators of positive and negative affect. Besides, participants also reported their subjective experience of social media use in the past hour. Specifically, they were asked to reflect on their use of the social media platform reported earlier and indicate how happy, meaningful, and effortful the platform made them feel. Responses captured momentary evaluations of platform-specific experiences.

Covariates

The baseline survey assessed individual characteristics related to digital well-being and social media use. Demographic information, such as age and gender, was collected via Prolific. The baseline survey also collected data on social ties (self-reported number of strong social ties), digital stress [52], problematic social media use [191], anxiety (GAD-2) [192], depression (PHQ-2) [193], social media mindset [53], and need for cognition [194]. The endline survey, administered between Days 7 and 10, reassessed social media mindset that includes measures of perceived agency and valence for social media use [53]. At each EMA prompt, we also measured the level of engagement to five distinct social media activities. Specifically, participants reported the number of minutes spent on social media activities during the past hour using slider scales ranging from 0 to 60 minutes. Activities included: (a) posting content, (b) browsing others' posts, (c) searching for specific information, (d) messaging, and (e) commenting on or liking others' content.

3.2.4 Ethics and transparency

The study procedures were approved by the Institutional Review Board (IRB) of the Harvard University-Area (IRB24-1356). All measures and procedures followed the study's preregistered protocol, which is publicly available at <https://osf.io/4rfs9/overview>. Data for analyses replication is available here.

3.2.5 Survey data analysis

A total of 266 U.S. adults participated in the EMA study, yielding 7,961 completed EMA surveys across the 7-day period, corresponding to an overall response rate of 71.26%. As specified in the power estimation in the pre-registration, assuming six prompts per day over seven days with a 50% response rate (~21 assessments per participant) and an intraclass correlation coefficient (ICC) of 0.40, a sample of 250 participants would provide over 90% power to detect small effects ($\beta = .05$ between-person; $\beta = .01$ within-person). As shown in Figure A2.1, survey

completion rates remained stable across study days, consistently exceeding 70%. Of the 7,961 EMA responses, 4,318 (54.2%) involved social media use in the preceding hour, whereas the remaining 3,643 responses reflected engagement in non-social media activities (e.g., learning, household tasks). The present analyses focus on the 4,318 EMA observations involving social media use. Descriptive analyses were conducted to characterize patterns of social media use and goal-related measures. All survey data processing and analyses were performed using Python 3.9 and R 4.2.1.

3.2.6 Screenshot analysis

Objective social media use time data were collected via participant-submitted screenshots of smartphone screen-time summaries. Text was extracted from screenshots using the same image-processing and text-recognition pipeline established in Chapter 2. Across the 7-day study period, participants were expected to submit 3,724 screenshots (two per day). A total of 2,452 screenshots were received, corresponding to a submission rate of 65.8%, shown in Figure A2.2. The screenshot submission rates remained stable and exceeded 60% across days, surpassing screenshot compliance rates reported in prior research using similar methods [99]. To ensure data validity, all submitted screenshots underwent a series of quality checks, motivated by prior evidence that participants may occasionally upload incorrect or irrelevant images. Three quality criteria were applied: (1) the timestamp of the screenshot had to align with the scheduled EMA prompt, (2) the screenshot had to display a daily (rather than weekly) screen-time view, and (3) the screenshot date had to match the corresponding survey date. Shown in Figure A2.3, of the submitted screenshots, 2,424 (95.4%) satisfied at least one quality criterion, and 2,046 (80.5%) satisfied all three criteria. Screenshots meeting all three quality checks were retained for analysis.

3.2.7 LLM-assisted content analysis on written description

We analyzed participants' open-ended responses to the question, "In a few words, can you describe why you opened [the social media platform reported used the most in the past hour] the last time you used it in the past hour?". Each response was classified into thematic categories representing distinct types of intentions for social media use. We developed a 10-category codebook, covering information seeking, social connection, self-expression, entertainment/relaxation, passing time, coping, habitual use, productivity/learning, trigger-driven use, and unspecified intention, with detailed descriptions and examples for each category. The codebook was initially developed based on prior literature on motivations for social media use [112,195]. The research team then randomly sampled 200 self-reported motivation responses, independently reviewed and categorized them, and iteratively refined the coding scheme through repeated discussion. This iterative process focused on distinguishing closely related constructs (e.g., passing time vs. coping, habitual vs. trigger-driven use) and ensuring that vague or purely activity-based responses were consistently classified as "Unspecified."

Classification was performed using GPT-4o, prompted with the full codebook and the survey response text. Each response was run through the model five times with temperature set to 0, and the final label(s) were selected based on the most probable output across runs. If no category aligned, the model returned "None." To evaluate reliability, we conducted manual double coding on a random sample of responses and compared GPT classifications with human coding using Cohen's kappa. This procedure provided a measure of model-human agreement while also allowing us to identify edge cases and further refine the codebook. The performance of the LLM-assisted text classification is presented in Table A2.1. The codebook of goal categories is presented in supplementary materials.

3.2.8 TikTok data analysis

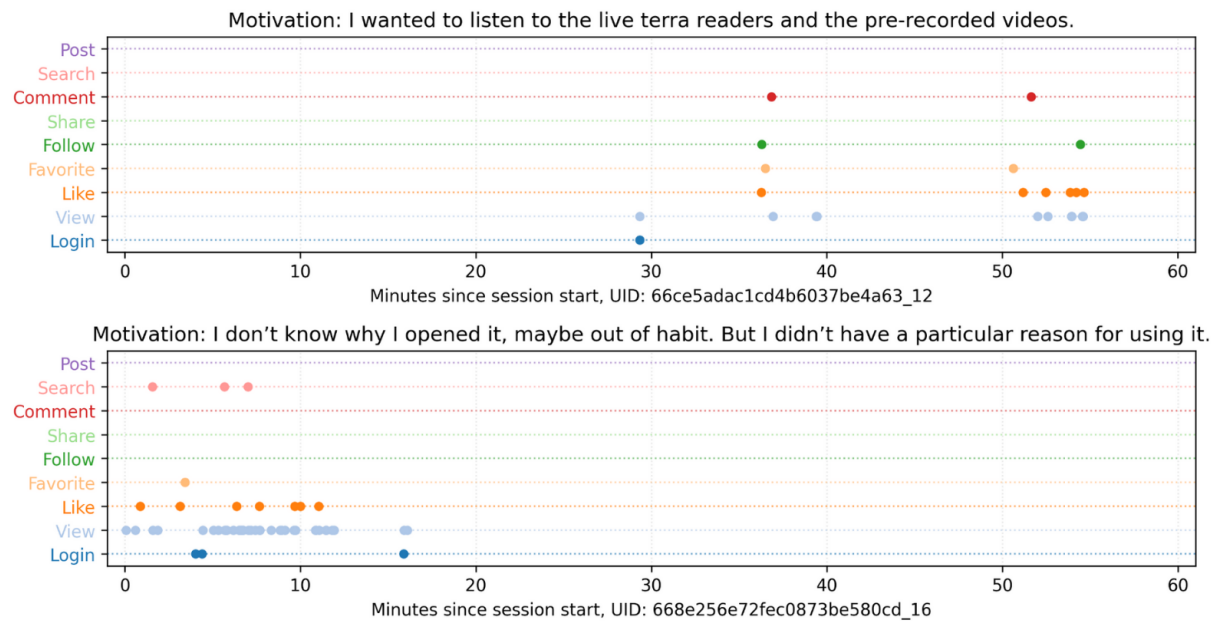
The donated data included 1) time-stamped records of TikTok activities, including viewing, posting, commenting, liking, sharing, following, searching, favoriting, and logging events; 2) URLs

of videos that were viewed, liked, shared, and marked as favorite; 3) text of the comments made by the user; 4) search terms the users input; 5) the usernames followed by the users. We observed 1918 EMA bursts with tiktok activities in the past 2 hours from 161 participants, resulting in 116020 activities. In the EMA data, because participants were asked to report their awareness of motivation for the most-used app in the past hour, we restricted analyses to EMA bursts in which TikTok was identified as the most-used app. This resulted in 1,234 EMA bursts from 180 participants. Among these, 1,000 bursts from 143 participants could be linked to TikTok use in the DDP objective data. The discrepancy may reflect (a) the fact that only 192 of the 266 participants shared their TikTok DDP logs, and/or (b) recall error or misreporting in the EMA self-reports.

Identify TikTok episodes matched with EMA survey

We first identified TikTok use episodes relevant to our research questions by linking the timing of TikTok use with the timing of EMA survey responses for each participant. Although many participants shared their complete TikTok use histories, our analyses focused on use sessions corresponding to the EMA recall window. Specifically, when a survey was responded at 10:00 a.m., participants were asked to report on their TikTok use during the preceding hour; accordingly, user responses to the motivation-related questions in the survey is based on their experience from 9:00–10:00 a.m, thus we linked survey responses to objective use data in that period. To account for potential recall imprecision regarding the one-hour window, we conducted a sensitivity analysis using an extended two-hour window (e.g., in this case, 8:00–10:00 a.m.). For each participant, we created episodes of TikTok use occurring within these matched time windows. The timing of different activities within the matched episodes is illustrated for each participant and each EMA survey, with examples shown in Figure 2.2.

Figure 2.2. The example time of TikTok activities in the episodes matched with EMA survey for two participants



Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Donated TikTok data included time-stamped records of user activities and content interactions, which were linked to EMA reports and restricted to bursts in which TikTok was identified as the most-used app. Data points represent a logged activity from user donated TikTok data.

Create indicators for basic activities

The TikTok usage data captures a rich and comprehensive record of users' experiences, including both user actions and the content they are exposed to. Therefore, there are numerous ways to unpack the data, that identify objective behavior signals and integrate the objective signals with subjective EMA self-reports to inform hypothesis testing. A comprehensive exploration of these possibilities, however, is beyond the scope of the current study. We note that, in a parallel project, we are working on systematically expanding TikTok data to examine whether users with different motivations exhibit distinct objective usage patterns, and whether the awareness or strength of these motivations shapes the TikTok experience. These ongoing projects investigate variation in multimodal video content exposure, including topic distributions,

semantic and sentiment characteristics, and functional video attributes (e.g., informational versus non-informational content).

In the present study, we focus on core behavioral indicators derived from TikTok use data, including (a) basic activities and (b) three constructed behavioral signals, including active use, session duration, and video skipping behaviors. Basic activities consist of specific TikTok actions reported in the user-shared data, including viewing, posting, commenting, liking, sharing, following, searching, favoriting, and logging events. For each activity type, we derived the following measures. These measures were computed separately for each activity type and used as inputs for subsequent analyses.

- Total activity counts in the matched use episodes: the cumulative number of occurrences of the activity in each use episode matched with the EMA survey, which is a within-person level variable.
- Total activity counts in entire use history: the cumulative number of occurrences of the activity across the participant's entire TikTok DDP history, which is a between-person level variable.
- Normalized activity rate: the total number of activities divided by the number of days since the participant began using TikTok, capturing average daily activity frequency, which is a between-person level variable.

Identify use sessions within TikTok use episode

A defining feature of TikTok use is that users frequently skip videos and engage with the platform in both continuous and discontinuous patterns. For example, a user may open TikTok once and scroll consecutively for five minutes, or open the app multiple times for brief one-minute sessions, sometimes prompted by push notifications. During continuous use, video-skipping behavior is particularly prevalent. Video skipping may indicate either mindless scrolling or active curation of one's content environment. Accordingly, we aim to examine these phenomena in

greater detail and to operationalize them using behavioral trace data. As a first step, we identify consecutive TikTok use sessions from the TikTok use data.

As illustrated in the bottom panel of Figure 2.2, from approximately minute 5 to minute 8, the user engages in consecutive video viewing, indicating continuous scrolling behavior, likely within the For You feed. Therefore, when TikTok video viewing events occurred consecutively, defined as the user initiating a new video before the previous one finished, we treated these viewing events as belonging to the same session. The length of a session was defined as the time difference between the start time of the first video and the end time of the last video in that sequence, where the end time was calculated as the start time of the last video plus its video duration. We defined sessions of activity at the level of each participant and each EMA survey window. For each viewing event, we computed an end time as the start time plus the video duration (obtained from the TikTok Research Tool, with missing durations imputed as 46 seconds, which is the mean video duration observed in the collected TikTok data). Viewing events were sorted chronologically, and consecutive events were chained into the same session if the gap between the end of one event and the start of the next was less than 60 seconds; otherwise, a new session was initiated. Each session was therefore characterized by a start time (the first viewing event), an end time (the end time of the final viewing event in the chain), and a session duration (end minus start). We further calculated the total number of sessions per participant per EMA survey burst, as well as the total observed duration of TikTok use as the sum of session durations for each TikTok use episode.

We note three additional analytical decisions made to enable session construction. First, although most non-viewing activities on TikTok occur alongside video viewing, because the app typically opens directly to the For You feed, there are instances in which users engage in activities such as searching or sharing without an identifiable viewing session. For these activities, use duration could not be directly observed, as only the event start time (but not the end time) was available in the TikTok data. In such cases, we imputed duration values based on activity type: 5

seconds for liking, following, and favoriting; 10 seconds for searching and sharing; 20 seconds for commenting; and 3 minutes for posting. Second, not all TikTok content includes explicit duration metadata. In particular, TikTok allows text- and photo-based posts, for which duration information is unavailable. We therefore imputed a duration of 10 seconds for text and photo posts and 46 seconds for video posts with missing duration information, where 46 seconds corresponds to the mean video duration in our dataset. Finally, due to data limitations, our active use duration measure cannot capture time spent repeatedly watching a single video when no subsequent observable activity occurs. In these cases, the duration of the final viewed video is truncated at its recorded length, leading to a potential underestimation of actual use time.

Create indicators for behavioral signals

Based on the identified sessions within each TikTok use episode, we constructed three higher-level behavioral signals from TikTok use data: observed use duration, active TikTok use, and video-skipping behavior. Observed use duration was operationalized as the total time participants spent on TikTok within each episode, calculated as the sum of session durations across all identified sessions.

Active TikTok use was constructed following the passive–active social media use literature [18] and captures user-initiated interactions beyond passive viewing, including liking, following, favoriting, searching, commenting, sharing, and posting. We operationalized active use using four indicators: (1) the total frequency of active behaviors within the matched episode; (2) the proportion of active behaviors relative to video views ($\# \text{ active behaviors} / \# \text{ views}$); (3) the number of active behaviors per second of actual TikTok use; and (4) the average frequency of active behaviors per active session, defined as the mean number of active behaviors per session within the observed period.

To capture video-skipping behavior, we constructed two sets of measures: (1) indicators of whether and how often videos were skipped, and (2) indicators of how much time participants

spent on a video before skipping. A video view was coded as skipped if the start time of the subsequent video occurred before the end time of the current video, where the end time was defined as the start time plus the video duration. We then computed the total frequency of video skips (Freq_vs). We further derived three normalized measures: the proportion of video skips relative to total video views (Ratio_vs), the number of video skips per second of observed TikTok use (Density_vs), and the average frequency of video skips per active session (Freqpersession_vs). In addition, for videos that were skipped, we calculated the time elapsed (in seconds) between the start of a video and the initiation of the subsequent video, representing the time spent on the video before skipping. For videos that were not skipped, this measure was coded as missing. To account for heterogeneity in video length, we also computed the ratio of time spent before skipping to the total duration of the video. Together, these measures capture both the frequency and temporal dynamics of video-skipping behavior during TikTok use episodes.

3.2.9 Descriptive data analysis

We conducted a series of descriptive analyses to characterize the sample, measured variables, and objective social media use. Demographic characteristics of the study sample are reported in Table A2.2. The distribution of the baseline problematic social media use is summarized in Figure A2.4, and the baseline social media mindsets are described through measures of agency (Figure A2.5) and valence (Figure A2.6). We next describe the distributions of the key independent variables, goal awareness and goal strength (Figure A2.7), as well as their variation across different motivation categories (Figures A2.8 and A2.9). Summary statistics for the dependent variables, independent variables, and covariates are reported in Table A2.3. Users' total screen time and screen time by app and app categories, extracted from user-shared screenshots, are summarized in Table A2.4. We also described objective TikTok use data, including overall usage patterns and three derived measures: observed use duration, active TikTok use, and video-skipping behavior (Table A2.5, Figure A2.10). We also examine two key

objective TikTok use variables, active activity per video view and the proportion of viewed videos that were skipped, across different self-reported TikTok use goals (Figure A2.11). As a sanity check for potential temporal patterns, we also examined whether time of day and day of week were associated with systematic differences in key variables, including goal awareness, goal strength, social media use experience, subjective well-being outcomes, social media activities (Figure A2.12), and TikTok objective use variables (Figure A2.13) by plotting their means and standard deviations across time.

3.2.10 Statistical analysis

We tested the study hypotheses using multilevel modeling to account for the nested structure of the EMA data, with repeated prompts nested within days and days nested within individuals.

Test Hypothesis 1

To test H1, which posits that higher goal awareness and goal strength in social media use is associated with more positive user experiences and better well-being outcomes at both the within-person and between-person levels, we estimated a series of three-level linear mixed-effects models (R, lme4) to account for the nested structure of repeated EMA observations within days and days within individuals. The primary independent variables of interest were goal awareness and goal strength, each operationalized at both the between-person (grand mean centered) and within-person (person mean centered) levels. Dependent variables included social media experience and subjective well-being.

All models additionally adjusted for a set of theoretically relevant covariates across multiple levels. Between-person covariates included age, gender, baseline need for cognition, baseline problematic social media use, baseline network social network size, baseline mindset

measures (agency and valence), and the most observed used app category (social media, game, browser, message) in the entire study period from screenshot. Day-level covariates included total time spent on social media during the day. Within-person covariates included the proportion of time spent browsing relative to other activities on social media use, and the platform use category of the past hour social media use session (TikTok, Facebook, Instagram, YouTube, Reddit, Twitter/X, message apps, and others). Random intercepts were specified for individuals and for survey days nested within individuals to account for unobserved heterogeneity at both levels.

To determine the appropriate random-effects structure, we estimated four candidate models for each outcome while holding the fixed-effects specification constant: (1) a random-intercept-only model, (2) a model with random intercept and random slopes for momentary goal awareness, (3) a model with random intercept and random slopes for momentary goal strength, and (4) a model with random slopes for both momentary goal awareness and goal strength (specified without correlated random effects). All models were estimated using maximum likelihood. Model fit was compared using Akaike Information Criterion (AIC), and the best-supported model for each outcome was selected based on the lowest AIC value among converged, non-singular models. Fixed-effect estimates were reported from the selected model. Results are presented in Table 2.1.

Test Hypothesis 2

To test H2, which posits that the associations between goal awareness, goal strength, and outcomes are consistent across different types of social media use goals (e.g., entertainment, information seeking, and social connection), we first estimated separate three-level multilevel models within subsets defined by goal type. The model specifications within each subset were identical to those used in the main analyses for H1. Results from these subgroup analyses are presented in Figure 2.3 and Table A2.6.

In addition, we tested H2 using a moderation framework by examining whether goal type moderated the within-person associations between goal awareness, goal strength, and outcomes at the momentary level. Specifically, we tested six types of social media use goals as moderators by estimating separate three-level multilevel models for each goal type. In each model, the outcome was regressed on the within-person and between-person components of goal awareness and goal strength, the moderator (goal type), and their interactions, along with the full set of covariates included in the H1 models. The focal test of moderation was the interaction between the within-person components of goal awareness or goal strength and goal type, which assessed whether the momentary associations varied as a function of social media use goal. Consistent with the H1 analyses, we evaluated multiple random-effects specifications (random intercepts and random slopes) and selected the final model based on the lowest AIC among non-singular models. Results of the moderation analyses are reported in Table A2.7.

Test hypothesis 3

For H3, we examined whether higher motivation awareness or goal strength was associated with specific objective tiktok use patterns. We conducted three sets of analyses. First, we examined both between-person and within-person correlations among TikTok objective use and goal awareness, strength, as well as social media experience and subjective well-being outcomes (Figure 2.4, Figure A2.14). To examine associations between repeated-measures variables, we estimated linear mixed-effects models with random intercepts for participants (Figure 2.4). Each model included a random intercept for participants to account for stable individual differences in the outcome across time. The independent variables, TikTok objective behaviors, were decomposed into within-person (person-mean centered) and between-person (grand-mean centered) components. Both the within-person and between-person components of the predictor were entered simultaneously as fixed effects in the model. The dependent variables are the goal awareness, strength, subjective well-being, and social media experience measured

in each EMA survey. To facilitate comparability of effect sizes across variables, standardized coefficients for both within- and between-person effects were derived post hoc by scaling the estimated coefficients by the ratio of the predictor's standard deviation to the outcome's standard deviation. As a robustness check, we additionally estimated between-person and within-person Pearson correlations (Figure A2.14). Between-person correlations were computed as Pearson correlations between person-level means of each variable, while within-person correlations were computed as Pearson correlations between person-mean-centered deviations across all observations. These correlation-based estimates capture the same decomposition of associations as the multilevel models.

Second, we tested whether objective TikTok use variables served as mediators of the association between the independent variables and the outcomes, thereby assessing potential indirect pathways (Table A2.8). We conducted multilevel mediation analyses using linear mixed-effects models to examine whether objective TikTok use mediated the within-person association between goal-related predictors (goal awareness and goal strength) and happy social media experience. Time was included as a covariate in both the mediator and outcome models.

Third, because the objective TikTok use data are high dimensional, the analyses described above may not adequately capture the complex and potentially nonlinear relationships between objective TikTok use and goal awareness, goal strength, and social media use experiences. To address this limitation, we employed random forest models to identify the most informative predictors of both dependent and independent variables (Figure A2.15). Specifically, we applied a Hybrid/Mundlak feature decomposition to the longitudinal, person-nested EMA data within a random forest framework, allowing us to disentangle within-person associations from between-person differences while retaining the flexibility to model nonlinear and high-dimensional relationships. Prior to model estimation, we winsorized selected predictor groups by replacing values above the 99th percentile with missing values to reduce the influence of extreme outliers. Missing values were then handled using mean imputation for numeric predictors and mode

imputation for non-numeric predictors within the analytic dataset. Random forest models were estimated using the ranger package with permutation-based variable importance, specifying 50 trees for importance extraction and setting mtry to $\lfloor p/3 \rfloor$, where p denotes the total number of predictors. Permutation importance ranks predictors by the increase in prediction error when that variable is permuted. To evaluate generalization under clustering, we used grouped 5-fold cross-validation by person ID, ensuring that all observations from a participant appear in only one fold. We report model performance using grouped cross-validated root mean squared error (RMSE), comparing the random forest predictions against a mean-only baseline to quantify improvement in out-of-sample prediction.

We plotted the top 20 most important variables with color-coded grouping to visualize whether importance is dominated by within-person signals vs between-person/stable differences. Specifically, we grouped predictors into conceptually distinct categories reflecting both measurement level and data source. The groups included time-varying independent variables (e.g., momentary goal awareness and goal strength) and covariates (e.g., type of goal, social media use activity) measured in the EMA survey at the within-person level (p_w); their between-person counterparts derived from person-level means across the EMA period (p_b); stable person-level characteristics such as demographics, anxiety, depression, social media use mindset, need for cognition, problematic social media use, total screen time, and dominant platform use (s_self). Objective TikTok use were incorporated at multiple levels, including long-term person-level activity aggregated across the full data donation period (s_ddp), person-level TikTok activity during the focal 14-day EMA window (ddp_b), and momentary within-person TikTok activity during the EMA period (ddp_w). Finally, we included social media experience and subjective well-being measures at both the between-person (wb_b) and within-person (wb_w) levels, as well as temporal context variables capturing day of week and time of day ($todw$).

Post hoc analyses on social media mindset

In addition to the preregistered tests of Hypotheses 1–3, we conducted a set of exploratory post hoc analyses informed by patterns observed in the data. We examined changes in agency and valence mindsets of social media use from baseline to post-EMA. The motivation for this analysis was that the EMA survey, where we repeatedly prompted participants to report their goal awareness and motivational strength, may itself function as an intervention, potentially shaping goal pursuit processes during social media use. Examining changes in the agency and valence mindset may help inform the design of future intervention studies. We therefore visualized the distribution of individual-level changes in agency and valence mindsets (endline minus baseline; Figure 2.4). To assess which baseline characteristics were associated with subsequent changes in agency and valence mindsets, we estimated two separate linear regression models predicting change in agency and change in valence, respectively. Both models included baseline demographic characteristics (age and gender), baseline mental health indicators (anxiety and depression), psychosocial traits (fear of missing out, social network size, and need for cognition), baseline social media use status (problematic social media use and baseline agency and valence mindsets), smartphone usage characteristics (total screen time), and indicators for primary app categories used (browser, gaming, messaging apps, and TikTok). Results are presented in Table A2.9.

3.3 Results

3.3.1 Descriptive Results

Baseline Characteristics

We first describe the characteristics of the study sample. As shown in Table A2.2, the average participant age was 38.1 years (SD=11.9 years). The sample was 61.5% female, with 60.9% identifying as White, 17.1% as Black, and 6.2% as Asian. In addition, 17.9% of participants were students, 4.1% were unemployed, and 5.3% were born outside the United States. In terms of the baseline characteristics of social media use, Figure A2.4 shows the distribution of

problematic social media use, measured on a scale ranging from 0 to 5. Approximately 15% of participants reported scores in the upper range of the scale (scores ≥ 4). Distributions of agency and valence mindsets are presented in Figures A2.5 and A2.6, respectively. For agency mindset, over half of participants reported scores above the midpoint of the scale, with a comparable proportion reporting scores below the midpoint, indicating substantial variability in perceived agency in social media use. In contrast, the distribution of valence mindset was skewed toward more positive evaluations, with over 75% of participants reporting values above the midpoint, reflecting generally positive attitudes toward their social media experiences.

EMA measures

Turning to the distribution of EMA measures and their variability, Table A2.3 summarizes the distributions of key EMA variables, including independent variables (goal awareness and strength), dependent variables (social media use experience and subjective well-being), and social media activity. On average, participants reported moderate levels of goal awareness ($M = 3.60$, $SD = 1.18$) on a 1–5 scale, with an intraclass correlation coefficient (ICC) of 0.52. Goal strength showed a similar pattern ($M = 3.36$, $SD = 1.16$; $ICC = 0.49$). These ICC values indicate that substantial variance in both measures was attributable to both between-person differences and within-person fluctuations over time.

For dependent variables, shown in Table A2.3, participants' momentary social media experiences in the past hour were assessed along three dimensions, happiness, meaningfulness, and effortfulness, each measured on a 1–5 scale. The mean level of happiness was 3.59 ($SD = 1.04$), meaningfulness averaged 3.41 ($SD = 1.22$), and effortfulness was lower on average ($M = 2.60$, $SD = 1.31$). ICCs for these experience measures ranged from 0.58 to 0.65, suggesting between-person differences accounted for a larger share of the variance than the within-person differences. Subjective well-being was assessed using a 0–100 scale. Average happiness was

61.13 (SD = 28.58), with an ICC of 0.75, indicating that a larger proportion of variability comes from the between-person differences than the within-level differences.

Finally, participants reported the amount of time spent engaging in different types of social media activities in the past hour. Browsing accounted for the greatest amount of time (M = 20.45 minutes, SD = 16.43; ICC = 0.50), followed by searching (M = 7.16 minutes) and commenting (M = 7.43 minutes), both of which exhibited ICCs exceeding 0.65. These patterns suggest that while browsing was the dominant activity overall, engagement in specific activities varied both across individuals and across time.

Distribution of goal awareness and strength

Figure A2.7 presents the distribution of goal awareness and goal strength across all EMA observations. Overall, substantial variability was observed at the person-time level in both the magnitude of goal awareness and the strength of motivation. Across EMA reports, 27.4% of observations indicated that participants were extremely aware of their goals, and 30.2% indicated high goal awareness. In contrast, 6.2% of observations reflected not being aware of the goal at all, with 12.0% reporting slight awareness and 24.1% reporting moderate awareness. A similarly wide distribution was observed for goal strength. Across EMA observations, 19.0% and 27.3% of reports indicated extreme or very high goal strength, respectively, while 31.0% reflected moderate strength. Lower levels of goal strength were less common, with 15.9% of observations reporting slight strength and 6.7% reporting not motivated at all.

Figure A2.8 further illustrates how goal awareness varied across self-reported goal categories. Goal awareness tended to be lower when social media use was motivated by entertainment (panel D), habitual use or passing time (panel E), or when no explicit goal was specified, such as when participants described only an activity or context (e.g., “scrolling in bed”) (panel F). In contrast, higher levels of goal awareness were more frequently reported when use was motivated by information seeking, social connection, self-expression, or external triggers

such as notifications (panel A-C). A similar pattern was observed for goal strength (Figure A2.9). Motivation tended to be weaker when social media use was driven by entertainment, coping or relaxation, habitual use, passing time, or unspecified goals, whereas stronger goal strength was reported when use was oriented toward information seeking, productivity, social connection, self-expression, or externally triggered engagement.

Screen time characteristics

Table A2.4 summarizes statistics on smartphone screen time derived from user-shared screenshots. Daily total screen time, measured in minutes, averaged 431 minutes (7.2 hours) per day (SD = 274), with a median of 386 minutes (6.4 hours). The intraclass correlation coefficient (ICC = 0.56) indicates that screen time exhibited substantial variability at both the between- and within-person levels. When disaggregated by app category (as defined by App Store classifications), social apps accounted for the largest share of screen time, with a mean of 166 minutes per day (median = 136 minutes; SD = 159). This was followed by productivity apps (mean = 57 minutes), entertainment apps (mean = 45 minutes), and gaming apps (mean = 34 minutes).

We further examined screen time at the app level. We combined Safari, Chrome, Edge, Firefox, and related applications as the browser app. Browser use accounted for the highest average app-level screen time (M = 71 minutes; ICC = 0.55), followed by TikTok (M = 52 minutes; ICC = 0.62). The long screentime of TikTok is partly due to the fact that the study recruited frequent TikTok users. Gaming apps accounted for an average of 30 minutes per day, followed by messaging apps (mean = 29 minutes), YouTube (mean = 25 minutes), Facebook (mean = 25 minutes), and Instagram (mean = 16 minutes).

We further note that these app category and app-level results should be interpreted with caution. The screenshots captured only the top three to five apps by usage, resulting in missing data for less frequently used applications. As a result, app-specific estimates likely reflect usage of the most heavily used apps rather than a complete account of total app engagement.

Objective TikTok use

We further describe TikTok-specific activities derived from user-donated TikTok data, which capture objective, real-time patterns of platform use at the individual level. As summarized in Table A2.5, participants spent an average of 27.1 minutes (SD = 22.3 minutes) using TikTok in the 2 hours preceding each EMA response, with an ICC of 0.33, indicating that a substantial proportion of variability occurred within individuals over time. Figure A2.10 illustrates the distribution of TikTok use over the two-hour observation window at the person-session level. As shown in panel A, over 65% of TikTok use sessions lasted less than 30 minutes, whereas approximately 8% lasted more than one hour. On average, participants engaged in four TikTok use sessions, defined as continuous periods of TikTok use separated by at least 60 seconds of inactivity, within the 2-hour observation window ($M = 4.2$, $SD = 3.4$). As shown in Figure A2.10 panel B, more than half of observation windows contained fewer than five TikTok sessions, whereas a small fraction (approximately 2%) contained more than 15 distinct sessions within the same two-hour time window.

We next examined the frequency of TikTok activities observed during the two hours preceding each EMA survey. Across observations, participants engaged in an average of 89 total activities (median = 65), including approximately 79 video views. In the 2-hour period, active behaviors were comparatively infrequent: participants clicked an average of four “likes,” conducted approximately one search, and engaged in fewer than one following (mean = 0.16 times), favoriting (mean = 0.38 times), sharing (mean = 0.31 times) or commenting (mean = 0.25 times) actions, with minimal content posting behaviors. Over 70% of two-hour observation windows involved fewer than 10 active behaviors (Figure A2.10 panel C), reflecting the predominantly browsing-oriented nature of TikTok use. To capture engagement intensity, we computed activity rates by normalizing activity counts by estimated minutes of TikTok use in the 2-hour time window. On average, participants viewed approximately 3.1 videos per minute, and

other actions such as liking, searching, or commenting occurred at rates well below one event per minute.

We additionally examined video-skipping behavior. Participants skipped an average of 59 videos during the two-hour observation window. Approximately 65% of viewed videos were skipped before completion. Shown in Figure A2.10 panel E, fewer than 35% of observation windows involved skipping fewer than 20 videos, while approximately 30% involved skipping between 20 and 60 videos. The distribution exhibited a pronounced right tail, with a small number of sessions involving skipping more than 300 videos, although this pattern was relatively rare. On average, participants skipped more than two videos per minute of use (Table A2.5). When examining the proportion of viewed videos that were skipped at the session level (Figure A2.10, panel F), approximately 20% of sessions involved skipping more than 80% of viewed videos, and roughly 50% involved skipping between 60% and 80%. In contrast, only a small fraction of sessions (approximately 3.8%) involved skipping very few or no videos during the observation window.

The mean time to video skip was 12.5 seconds, and the ratio of time watched to total video duration averaged 0.25, suggesting that participants typically viewed only the first quarter of a video before moving on. Shown in Figure A2.10, panel G and H, over 60% of videos were skipped within the first 10 seconds of playback, with a small proportion exhibiting extremely rapid skipping (within 5 seconds); more than 70% of videos were skipped before 30% of the total video duration had elapsed.

Finally, we examined the objective TikTok active engagement and video-skipping behavior across self-reported TikTok use goal categories (Figure A2.11). Overall, there were no substantial differences in the frequency of active engagement behaviors across goal groups. However, as shown in Figure A2.11 panel B, trigger-driven TikTok use (e.g., use initiated by receiving a notification) was associated with a lower proportion of skipped videos compared with use motivated by habitual engagement, passing time, entertainment, or coping. Use motivated by

social connection or self-expression also tended to be associated with a higher proportion of skipped videos relative to trigger-driven use. These differences were modest and did not show statistical significance, and thus should be interpreted as descriptive patterns with cautious.

Temporal variations

Finally, as shown in Figure A2.12, we did not observe systematic temporal differences in self-reported social media experiences or goal-related variables, including goal awareness and goal strength, nor in subjective well-being outcomes across the day. In contrast, Figure A2.13 illustrates greater temporal variability in objective TikTok use measures. Although these patterns did not reach statistical significance, descriptive trends suggest that video-skipping behavior tended to be slightly higher during evening and later hours of the day. Similarly, both the frequency of TikTok use and total observed use duration appeared to increase later in the day.

3.3.2 Results of H1

Table 2.1 presents the within- and between-person associations between social media goal awareness, goal strength, and two sets of outcomes: social media use experiences and subjective well-being. At the within-person level, social media use characterized by higher goal awareness was associated with greater happiness (standardized coefficient, $\beta = 0.04$, $SE = 0.01$) and lower loneliness ($\beta = 0.02$, $SE = 0.01$). Higher goal awareness was also associated with more positive momentary social media experiences, including feeling happier ($\beta = 0.05$, $SE = 0.01$), perceiving the experience as more meaningful ($\beta = 0.06$, $SE = 0.01$), and experiencing social media use as more effortful ($\beta = 0.03$, $SE = 0.01$). Similarly, at the within-person level, greater goal strength was positively associated with higher general happiness ($\beta = 0.05$, $SE = 0.01$) and feeling more inspired ($\beta = 0.06$, $SE = 0.01$). Goal strength was also strongly related to momentary social media experiences, including greater happiness ($\beta = 0.16$, $SE = 0.02$), greater meaningfulness ($\beta = 0.21$, $SE = 0.02$), and higher effortfulness ($\beta = 0.04$, $SE = 0.01$).

At the between-person level, goal awareness was not significantly associated with the outcomes. In contrast, individuals who, on average, reported stronger goal strength tended to experience higher levels of happiness ($\beta = 0.35$, $SE = 0.08$) and inspiration ($\beta = 0.34$, $SE = 0.07$), as well as more social media experiences with greater happiness ($\beta = 0.47$, $SE = 0.06$), meaningfulness ($\beta = 0.45$, $SE = 0.05$), and effortfulness ($\beta = 0.23$, $SE = 0.08$).

Table 2.1. Between-person and within-person associations between social media motivation awareness, strength, use experience, and well-being

	Subjective well-being				SMU experience was		
	Happiness	Inspired	Less Distressed	Less Lonely	Happy	Meaningful	Effortful
Goal awareness (between)	0.12 (0.08)	0.01 (0.07)	0.12 (0.08)	0.00 (0.09)	-0.01 (0.06)	0.04 (0.05)	-0.06 (0.08)
Goal awareness (within)	0.04 (0.01)*	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)*	0.05 (0.01)*	0.06 (0.01)*	0.03 (0.01)*
Goal strength (between)	0.35 (0.08)*	0.34 (0.07)*	-0.09 (0.08)	0.13 (0.09)	0.47 (0.06)*	0.23 (0.05)*	0.23 (0.08)*
Goal strength (within)	0.05 (0.01)*	0.06 (0.01)*	0.01 (0.01)	0.02 (0.01)	0.16 (0.02)*	0.04 (0.02)*	0.04 (0.01)*
VPC (person)	0.66	0.64	0.53	0.69	0.38	0.34	0.59
VPC (day)	0.09	0.07	0.17	0.09	0.07	0.06	0.04

Note: Data comes from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Reported results are derived from three-level linear mixed-effects models that account for repeated EMA observations nested within days and individuals and adjust for relevant covariates. * indicates P value < 0.05. The VPC values indicate the proportion of total outcome variance attributable to stable differences between individuals (person-level), day-to-day fluctuations within the same individual (day-level), and moment-to-moment variability within a day (residual level), respectively.

Taken together, these findings highlight the role of goal awareness in shaping social media experiences at the within-person level: when individuals use social media with greater awareness of why they are engaging in a given moment, they report happier and more meaningful social media experiences and better well-being outcomes. In addition, goal strength is associated with social media experiences and well-being at both the within- and between-person levels, indicating

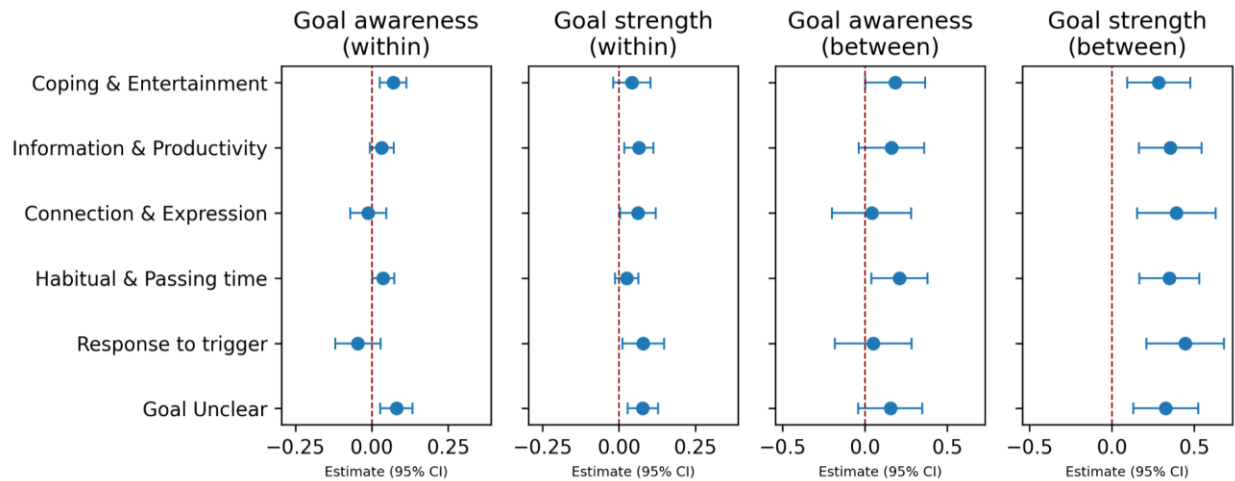
that users who are more strongly motivated at a particular moment, as well as those who are generally more motivated overall, have higher chances of reporting happier and more meaningful social media experience and well-being outcomes. These patterns are consistent with our hypotheses.

Interestingly, greater goal awareness and goal strength were associated with social media use experiences that were perceived as more effortful, yet also happier and more meaningful. This pattern sheds light to the discussion around balancing users' immediate need satisfaction and long-term value-aligned engagement. We will return to this point in the Discussion.

3.3.3 Results of H2

We further examined whether the associations between goal awareness, goal strength, and study outcomes varied across subgroups defined by self-reported social media use goals in the EMA surveys. Results from subgroup analyses are presented in Table A2.6, with Figure 2.3 providing a visual summary of associations for the happiness outcome. Across goal types characterized by coping or entertainment motives ($\beta = 0.07$, $SE = 0.02$), habitual use or passing time ($\beta = 0.04$, $SE = 0.02$), as well as instances in which no explicit goal was specified ($\beta = 0.08$, $SE = 0.03$), goal awareness at the within-person level was consistently and positively associated with momentary happiness (Figure 2.3, Table A2.6). In addition, at the between-person level, individuals who reported higher goal awareness during habitual or passing-time use ($\beta = 0.21$, $SE = 0.09$) or uses for coping or entertainment motives ($\beta = 0.18$, $SE = 0.09$) also reported higher overall levels of happiness. These patterns indicate that goal awareness is associated with greater well-being even in contexts typically characterized by low conscious awareness of goals or intentionality.

Figure 2.3. Associations of goal awareness and goal strength with happiness, by subgroup of momentary goal types



Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Reported results are derived from separate three-level linear mixed-effects models within subsets defined by social media use goal type, using identical model specifications to those in the main analyses in H1.

Turning to goal strength, within-person analyses showed that when social media use was motivated by information seeking or productivity ($\beta = 0.07$, $SE = 0.02$), social connection or self-expression ($\beta = 0.06$, $SE = 0.03$), externally triggered engagement ($\beta = 0.08$, $SE = 0.03$), or when goals were unspecified ($\beta = 0.08$, $SE = 0.03$), higher momentary goal strength was significantly associated with greater happiness (Figure 2.3, Table A2.6). This positive association between goal strength and happiness was also observed at the between-person level across all goal categories, indicating a robust relationship between stronger motivational engagement and well-being.

Across outcomes (Table A2.6), at the within-person level, higher goal awareness was consistently associated with more positive social media experiences, particularly greater happiness and meaningfulness, among use episodes characterized by coping or entertainment goals, as well as habitual or passing-time use. Similarly, higher goal strength, at both the within- and between-person levels, was positively associated with perceiving social media experiences

as happier and more meaningful when use was driven by information seeking or productivity, coping or entertainment, habitual or passing-time use, and when goals were not clearly specified.

We further conducted moderation analysis to test H2 to complement the subgroup analysis. Shown in Table A2.7, we examined whether the associations between social media use, goal awareness, goal strength, and happiness were moderated by goal-type subgroups. Overall, we found no strong evidence of moderation. The only exception is that, when social media use was motivated by social connection or self-expression, the within-person association between goal awareness and happiness appeared weaker compared with instances in which social media use was not driven by connection or expression motives.

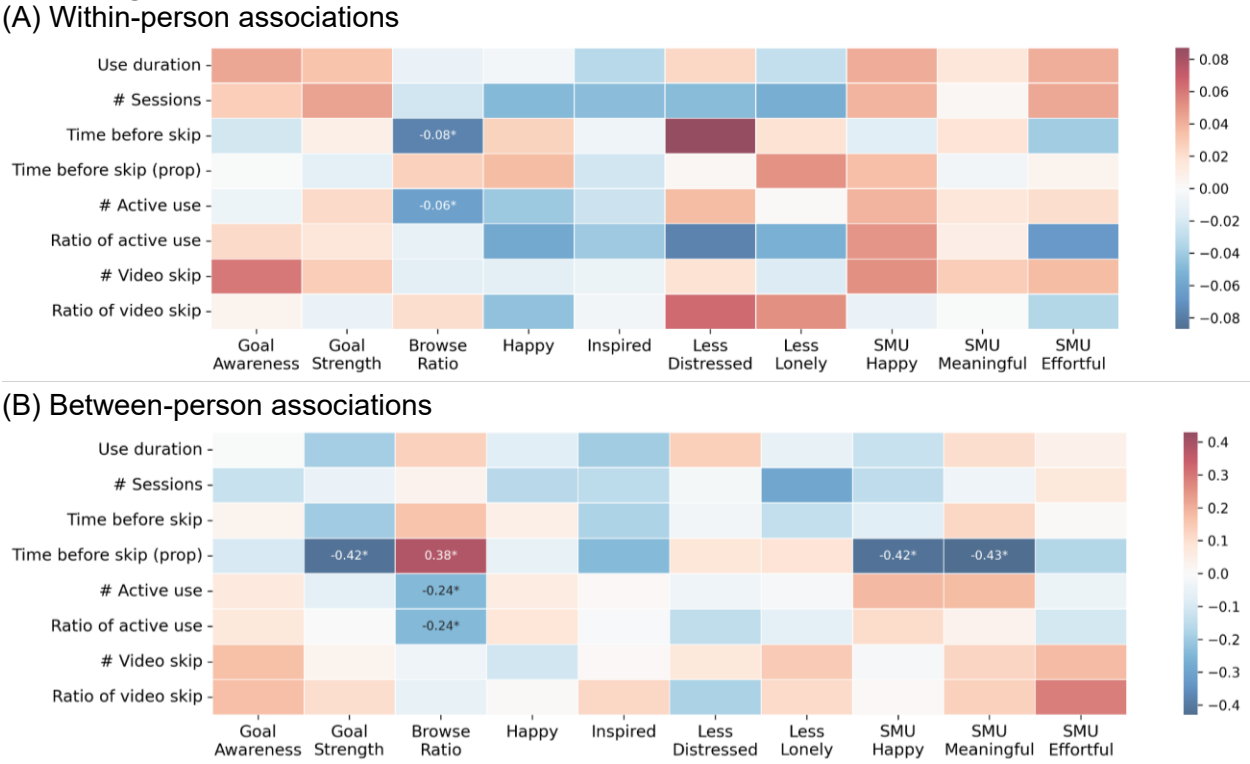
In summary, the results indicate that the relationships between goal awareness, goal strength, and the study outcomes are not generally moderated by goal type. At the within-person level, higher goal awareness was associated with more positive experiences and better well-being outcomes when the goals of using social media are entertainment or coping, habitual or passing-time use, and unspecified; higher goal strength was associated with more positive experiences and better well-being outcomes when social media is used for information seeking or productivity, social connection or self-expression, or externally triggered engagement. At the between-person level, the positive association between goal strength and well-being remained consistent across goal categories. Together, these findings support H2.

3.3.4 Results of H3

We next examined the within- and between-person associations between objective TikTok use behaviors, including duration of use, number of sessions, video-skipping behavior, and active engagement, and independent variables (goal awareness, goal strength), as well as the outcomes (social media experiences, and subjective well-being). The goal is to examine whether individuals who use social media with greater goal awareness or goal strength engage with TikTok in objectively different ways. More broadly, we examine whether measurable, objective behavioral

signals correspond to the psychological constructs we assessed. As shown in Figure 2.4 and Figure A2.14, we found little evidence of systematic associations between the extracted objective TikTok behaviors and goal awareness or goal strength, or with the study outcomes. Most pairwise associations did not reach statistical significance at the $p < 0.05$ level, and these analyses were not adjusted for multiple comparisons.

Figure 2.4. Within-person (A) and between-person (B) associations between TikTok use behaviors and goal awareness, goal strength, social media use experience, and subjective well-being



Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Donated TikTok data included time-stamped records of user activities and content interactions, which were linked to EMA reports and restricted to bursts in which TikTok was identified as the most-used app. Estimates are from linear mixed-effects models with participant-level random intercepts, decomposing objective TikTok behaviors into within-person (person-mean centered) and between-person (grand-mean centered) components and reporting standardized coefficients. A robustness check was performed with Pearson correlation analysis, shown in Figure A14.

Despite the largely nonsignificant results, several descriptive patterns are noteworthy. At the within-person level, a higher proportion of self-reported time spent browsing (relative to other

TikTok activities) was associated with shorter viewing durations before videos were skipped, indicating faster skipping in sessions dominated by browsing. Sessions characterized by higher levels of browsing also tended to exhibit lower levels of active engagement behaviors. In contrast, at the between-person level, individuals who spent a higher proportion of time watching videos before skipping, indicating longer viewing relative to total video duration, tended to report slightly lower levels of happiness, lower perceived meaningfulness of social media use, and lower goal strength, as well as a higher proportion of time spent browsing.

In the multilevel mediation analysis (Table A2.8), goal awareness was positively associated with happy social media experience ($dy:X = 0.21$, $SE = 0.04$, $p < .05$), but was not significantly related to the number of active TikTok activities ($dm:X = -0.02$, $SE = 0.34$), nor was the mediator associated with happy experience ($dy:M \approx 0$). This pattern was consistent across multiple objective TikTok use measures and outcome variables, as well as for both goal awareness and goal strength, with goal-related predictors showing robust direct associations with social media experience but little evidence that objective TikTok behaviors served as mediators. Accordingly, there was no evidence that the objective TikTok use behaviors we created mediated the association between goal awareness, goal strength and social media experience, subjective well-being.

Figure A2.15 presents the results of the hybrid random forest analyses across four outcomes. In Panel A, where the outcome is an objective TikTok behavior measured as raw time before video skip, 13 of the top 20 predictors reflect momentary, within-person TikTok activity, consistent with the fact that both predictors and outcome are derived from the same data source of TikTok objective data. Notably, within-person happiness and subjective well-being rank as the second- and third-most important predictors, suggesting that momentary affective states are associated with video skip patterns on the platform. Panel B shows results for goal awareness of social media use, where the strongest predictors are trait- and state-level goal strength, followed by trait-level subjective well-being. Panel C, which examines goal strength as the outcome,

reveals a similar pattern to Panel B. Panel D focuses on happiness as the outcome, after excluding other well-being and experience variables due to high correlations, trait-level goal awareness and goal strength remain the most important predictors, and the number of videos skipped per episode—also shows a modest level of importance. The observed RMSE reductions are generally modest, ranging from 0.1 to 0.2 standard deviations relative to the baseline, we interpret the variable importance results with caution. Overall, these findings indicate that objective TikTok interaction patterns, particularly video skipping behavior, are related to subjective well-being, although these associations should be interpreted cautiously given the exploratory nature of permutation-based variable importance.

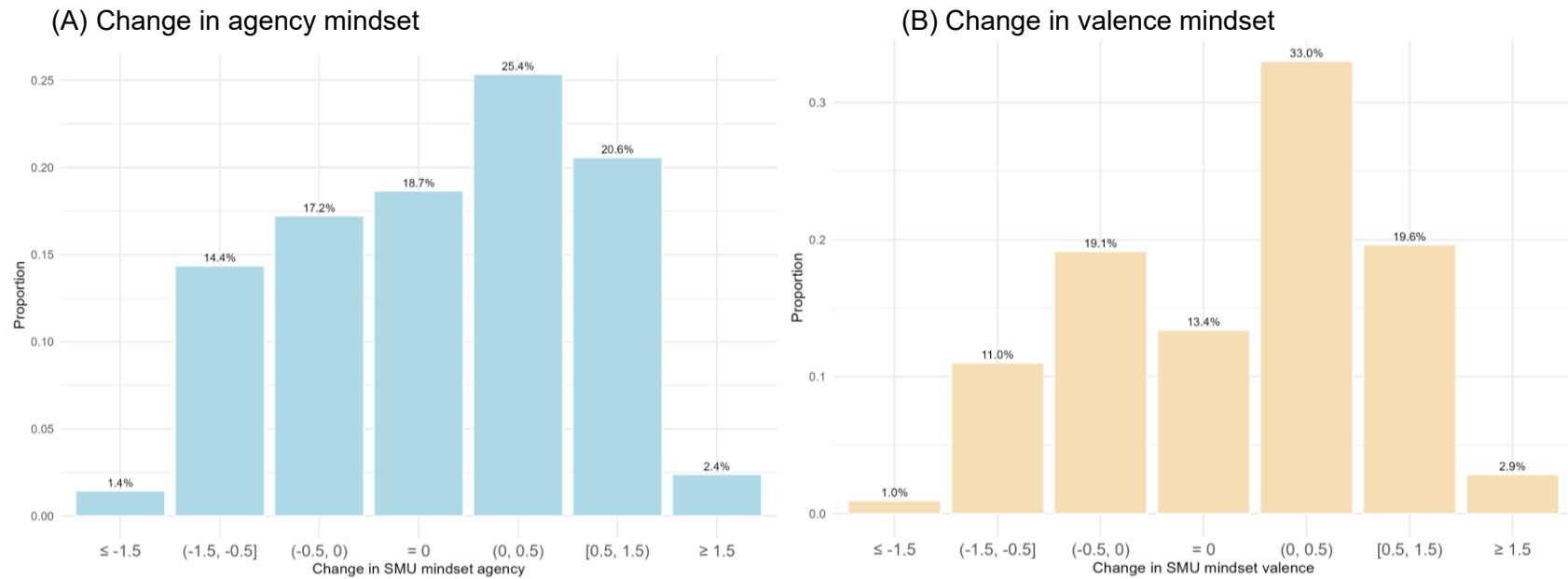
3.3.5 Change in social media use agency

As shown in Figure 2.5, approximately 48% of participants reported an increase in agency of social media use (feeling more controlled for social media use), while fewer than 35% experienced a decrease. Similarly, over 55% of participants reported increases in valence (social media use experience in positive), whereas fewer than 30% experienced a decrease in valence.

We next examined which baseline characteristics were associated with greater improvements in agency and valence (Table A2.9). For changes in agency, higher baseline agency was associated with smaller subsequent increases ($\beta = -0.53$, $SE = 0.02$), suggesting a ceiling effect, whereas higher baseline valence ($\beta = 0.04$, $SE = 0.02$) was associated with greater increases in agency over time. Several other characteristics were also associated with greater increases in agency, including longer average smartphone screen time ($\beta = 0.09$, $SE = 0.01$) and reporting TikTok ($\beta = 0.08$, $SE = 0.02$), browsing ($\beta = 0.10$, $SE = 0.02$), gaming ($\beta = 0.13$, $SE = 0.02$), or messaging apps ($\beta = 0.07$, $SE = 0.02$) as the most frequently used app category, relative to other app types. A broadly similar pattern emerged for changes in valence. Participants with lower levels of baseline valence ($\beta = -0.57$, $SE = 0.02$) were associated with smaller subsequent increases in valence. Participants who spent more time on their smartphones and those who

primarily used browsing, gaming, messaging apps, or TikTok as their most frequently used app category exhibited greater increases in valence over time. Together, these findings suggest that both baseline mindsets and patterns of smartphone and app use are associated with changes in agency and valence following the EMA period.

Figure 2.5. Distribution of changes in social media use agency (A) and valence (B) mindsets



Note: Agency and valence were measured on a 5-point scale (range = 1–5) at both baseline and endline. Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent.

3.4 Discussion

This study conceptualizes social media use as momentary goal pursuit and examines this process using a seven-day observational EMA design that integrates self-reports with participant-shared smartphone screenshots of screentime and, more importantly, objective TikTok data from data donation package. At the within-person (state) level, we find that users who were more aware of why they were engaging with social media, or who felt more strongly motivated at the moment of use, reported happier, more meaningful, and more effortful social media experiences, as well as higher levels of subjective well-being. Importantly, these associations were broadly robust across different types of social media use goals, suggesting that the benefits of goal awareness and goal strength are not confined to specific goals of use. In contrast, objective indicators of TikTok use showed limited and inconsistent associations with goal-related constructs and well-being, underscoring the challenges of inferring users' subjective experiences from high-dimensional and context-dependent behavioral traces alone. Although we observed some signals suggestive of nonlinear associations between video-skipping behavior and happiness, these findings should be interpreted with caution. In the sections below, we discuss the findings from the perspectives:

3.4.1 Goal awareness reflection as a user-centered digital intervention

A central implication of our findings concerns the potential of momentary goal awareness as a user-centered digital literacy intervention. Across moments of social media use, individuals reported more positive experiences and better subjective well-being when they were more aware of why they were engaging with social media at that moment. This state-level association suggests that interventions that prompt users to reflect on their goal before or during use may meaningfully improve digital well-being. This interpretation aligns with widely shared intuitions from users. Anecdotal reports indicate that many users desire simple reminders or mindful pauses that help them reconnect with their purpose for using social media and reduce distraction [196].

Such intuitions have motivated prior work in psychology and human–computer interaction [197–200], as well as the development of consumer-facing tools designed to introduce brief friction before app use [201–203]. For example, the One Sec app has a built-in intervention that prompts users to reflect on their goals before opening an app [202]. The intention reflection intervention supports more mindful engagement by making users’ motivations salient at the moment of action. Our findings provide empirical support for this approach by showing that goal awareness, even when measured repeatedly in everyday contexts, is associated with better experiential and well-being outcomes.

Notably, although not preregistered, we observed exploratory evidence that more than half of participants showed improvements in their social media use mindsets, including perceived agency and valence, over the seven-day EMA period, with larger improvements observed among participants with higher overall screen time. This pattern suggests that repeatedly prompting users to reflect on their motivations and experiences may itself function as a low-intensity intervention, strengthening users’ sense of control and fostering more positive appraisals of social media use. Building on these findings, we have started an experimental study that examines whether intention-reflection interventions embedded in everyday use contexts can causally shape user experiences and well-being outcomes.

3.4.2 Applying goal pursuit theory in social media use

Another contribution of this study is its conceptualization of social media use as a form of momentary goal pursuit. Our work applies psychological goal theories to explain users’ experiences across diverse types of social media engagement. By treating social media use as a sequence of goal-directed episodes rather than a monolithic behavior, we show that goal awareness and goal strength are related to users’ experiences across a wide range of goals for use. In this section, we discuss the lessons learned and unresolved challenges that arise from applying the goal pursuit framework to social media use.

One initial concern motivating this work was whether goal awareness would meaningfully vary in social media contexts, given that much social media use is characterized as habitual and automatic. Contrary to this concern, our data shows substantial moment-to-moment variation in both goal awareness and goal strength, indicating that habitual use does not preclude conscious awareness of goals during social media engagement. Moreover, the distribution of goal awareness varied by goal type in conceptually coherent ways: when goals were concrete, such as information seeking, social connection, or responding to notifications, users reported higher levels of goal awareness, whereas use motivated by entertainment, passing time, or habit showed lower and more dispersed levels of awareness. These patterns suggest that social media use encompasses both highly intentional and weakly intentional episodes.

A second puzzle concerned whether the benefits of goal awareness and goal strength would be limited to particular “constructive” goals, such as learning or social connection, rather than extending to more hedonic or habitual forms of use. Instead, our findings suggest the opposite: goal awareness plays a particularly important role in predicting social media use experiences in use episodes where goal awareness may be low, such as coping, entertainment, passing time, and habitual use. In contrast, for goal contexts where users already reported high levels of awareness (e.g., information seeking or social connection), associations between goal awareness and social media use and well-being outcomes were weaker or nonsignificant, likely reflecting ceiling effects and reduced variability in awareness. Together, these results suggest that it is not the content of users’ goals per se, but the degree to which those goals are consciously accessible and motivationally engaged at the moment of use, that shapes social media experiences.

Despite these advances, two unresolved questions remain. First, goal awareness and goal strength were moderately to strongly correlated in our data (Figure A2.16), raising important questions about cases in which users report high awareness of a goal that is nevertheless weakly motivating with low strength. Although we conducted exploratory analyses using participants’

open-ended descriptions of their motivations, we did not identify clear qualitative patterns that distinguished such cases. Understanding when and why users are consciously aware of weak or minimally motivating goals, and how these states differ from both strong goal pursuit and mindless engagement, remains an open question for future research.

A second unresolved puzzle concerns how habitual social media use should be conceptualized within a goal-pursuit framework. In highly automatized behaviors, analogous to fastening a seatbelt upon entering a car without the need to consciously think about road safety, individuals may act with minimal or no conscious awareness of underlying goals. Yet EMA surveys necessarily prompt users to reflect on their goals. For social media use episodes that are highly habitual, the survey may potentially induce post hoc inference of ones' goals. In other words, we do not know the cognitive processes users engage in when asked to report goals during highly habitual use, and whether such reports reflect genuine momentary awareness or reconstructed explanations. Addressing the challenge of integrating habitual use into theories of goal-directed social media behavior is essential for future work aimed at improving digital literacy, promoting mindful use, and informing product design and policy that center user values [57]. It is also critical for developing measurement approaches that more precisely distinguish automatic engagement from reflective goal pursuit.

3.4.3 Designing for user values and agency

Beyond implications for user psychology, our findings also speak to broader questions about how goal pursuit and agency can be incorporated into the design of digital platforms. A large body of literature characterizes social media use as a domain in which users struggle with hyperbolic discounting [62,155,204], facing a persistent tension between short-term gratification and longer-term value-based use goals, which further indicate failures in users' self-control [205]. Analogous to dietary choices in which immediate pleasure often overrides long-term health goals, social media platforms are frequently framed as environments optimized to deliver rapid, highly

salient rewards that undermine users' self-regulation over time. From this perspective, momentary social media use is often implicitly treated as inherently misaligned with users' longer-term interests.

Our results complicate this narrative by showing substantial heterogeneity in users' momentary goals and experiences. At the state level, users reported wide variation in what they sought from social media at each moment of use, and greater goal awareness was associated not only with increased happiness, but also with greater perceived meaningfulness and effort of the use episodes. These findings suggest that momentary engagement need not be reduced to impulsive or low-effort consumption driven solely by immediate reward. Instead, even within short time horizons, social media use can reflect intentional, value-aligned goal pursuit that is experienced as both effortful and meaningful. Our findings suggest a possible role for momentary goal awareness in supporting social media use that balances immediate needs with broader values and agency.

Although our analyses focus on users' momentary goals and experiences, our findings should not be interpreted as placing responsibility for digital well-being solely on individual users [21]. Rather, they point to an important role for platform design in supporting users' goals and agency. When users are more aware of their goals or feel more strongly motivated at the moment of engagement, they report better experiences and well-being in our study. This finding suggests that platforms could meaningfully improve user outcomes by designing environments that detect, respect, and align with users' goals. From this perspective, fostering intentional engagement is not merely a matter of individual self-control, but a shared responsibility in which platforms shape the choice architectures and affordances that enable goal-consistent use. The third chapter of the dissertation contains more details around this point.

At the same time, our findings highlight a fundamental design challenge: users' goals are dynamic, context-dependent, and heterogeneous across both moments and individuals. Goals may vary at the state and trait levels, shift rapidly within sessions, and coexist in ways that are

difficult to infer from behavioral signals alone. Designing systems that meaningfully respond to such complexity, while allowing users to articulate, reflect on, and revise their goals, remains an open problem. Despite the challenge, the bottom line is that rather than optimizing solely for engagement, platforms need to develop mechanisms that help users clarify what they want in a given moment and curate experiences accordingly. While addressing these challenges is beyond the scope of the present study, our findings motivate future work on how platforms can operationalize goal-sensitive design in ways that enhance user agency without overburdening users.

Finally, we note that these implications extend beyond social media platforms to the emerging design of LLMs, AI chatbots, and the general AI-driven tools. Much like early social media, contemporary AI systems are primarily developed and evaluated through technical benchmarks emphasizing accuracy, efficiency, or task performance [206]. Yet these systems increasingly function as interactive products embedded in users' everyday lives, shaping experiences, decisions, and well-being. Our findings underscore the importance of incorporating user goals and values into the evaluation and design of such systems. As AI technologies continue to evolve, insights from social science, human-computer interaction, and user experience research will be essential for ensuring that these systems align with users' intentions rather than exploit their attention, attachments, or sense of agency.

3.4.4 Implication from the analysis on objective TikTok

In this final section, we discuss findings on the associations between objective TikTok use behaviors and goal awareness, goal strength, and well-being. We initially hypothesized that individuals who engage with social media with greater goal awareness or stronger motivation would exhibit systematically different patterns of TikTok use compared to those with lower levels of goal pursuit processes. Contrary to this expectation, we found little evidence of consistent associations between objective TikTok behavioral indicators and goal-related psychological

constructs or social media use experience and well-being outcomes. While a small number of descriptive patterns emerged, such as associations between faster video skipping and momentary happiness, these results should be interpreted cautiously given the lack of consistent findings.

One interpretation of the lack of behavioral signals about goal awareness from objective TikTok data is that our current analytic approach may be insufficient to capture the high-dimensional and context-dependent nature of TikTok use. The same observable behaviors, such as video skips, can reflect qualitatively different psychological states. For example, rapid video skipping may indicate disengaged, mindless browsing in some contexts, but deliberate exploration or efficient filtering in others; conversely, prolonged viewing may reflect either deep interest or passive immersion driven by algorithmic recommendation. As a result, simple behavioral summaries, such as average viewing duration or skipping rates, may be insufficient to capture the underlying psychological meaning of platform use. From this perspective, the absence of robust behavioral signals for goal pursuit process from objective TikTok use data may reflect a mismatch between the complexity of users' experiences and the granularity of analyses we conducted.

Alternatively, the divergence between self-report and behavioral associations may also reflect common-method bias [207]. Because goal awareness, goal strength, and well-being are all assessed via the same self-report EMA instrument, while objective TikTok behavioral indicators are measured in a different way. Shared method variance, such as acquiescence bias and social desirability, may partially inflate their inter-associations, while objective TikTok behaviors are not subject to the same inflation. Although the real-time, in-the-moment nature of EMA assessments reduces several traditional sources of common-method bias relative to cross-sectional, retrospective surveys, we acknowledge that it does not eliminate them entirely.

The mismatch between the complexity of social media use experience and analytic granularity points to important directions for future research. Rather than focusing solely on

aggregated behavioral metrics (e.g., active engagement or video-skipping rates), future research may benefit from incorporating the content and contextual features of social media use into the analysis. One promising hypothesis is that intentional social media use may be reflected less in how users interact with content (e.g., how quickly they skip) and more in what content they are exposed to and engage with. For instance, individuals with higher goal awareness may encounter or select more homogeneous, goal-aligned content that reflects stable interests or higher cognitive engagement, whereas less intentional use may be associated with more heterogeneous or repetitive content optimized primarily for attention capture. Examining content characteristics alongside behavioral traces may therefore provide a more sensitive lens for detecting how goal pursuit manifests in digital environments.

An alternative interpretation of the largely null associations between objective TikTok use behaviors and goal awareness, goal strength, and well-being is that subjective experience and objective behavior may be only weakly related. Under this account, individuals' self-reported awareness, motivation, and well-being reflect subjective appraisals of their engagement that are not tightly mapped onto observable behavioral traces. Although such a dissociation may appear counterintuitive, similar discrepancies between subjective judgments and objective behavior have been documented in social media use [208], as well as across multiple domains in psychology, such as memory [209] and attention [210,211].

If subjective experience and objective behavior are indeed partially decoupled, the implications for digital well-being shift accordingly. Rather than focusing exclusively on modifying behavioral patterns or usage metrics, interventions may be more effective when they target users' interpretations, appraisals, and sense of agency regarding their own behavior. This perspective aligns closely with emerging work on social media mindsets [53] and agency in social media use. From this standpoint, digital well-being may be less about achieving externally defined "optimal" usage patterns and more about supporting users in developing self-understandings and

experiences of use that feel intentional, agentic, aligned with their goals, and psychologically satisfying.

3.4.5 Limitations

Several limitations of this study should be acknowledged. First, our design relies on participants' self-reports of goal awareness and motivational strength over the past hour. These measures are metacognitive in nature and inherently susceptible to self-perception processes [212]. In the context of EMA, participants may infer or reconstruct their motivations post hoc rather than report directly accessible mental states. Such reports may therefore be influenced by recall bias, social desirability bias, or broader self-appraisal tendencies, and may not perfectly reflect momentary motivational states as they unfolded in real time. This concern may be particularly salient in contexts of habitual or automatic use, where goals may not be fully accessible to conscious awareness, as discussed above. Moreover, given the observational nature of the study, we cannot fully rule out confounding between well-being outcomes and biases from the self perception process. For example, individuals experiencing more positive momentary affect may be more likely to construe their engagement as intentional or goal-directed, simultaneously reporting higher goal awareness, stronger motivation, and better well-being. To address these concerns, we are conducting a follow-up study employing an experimental pop-up survey design that prompts users to report goal awareness at the moment of app opening. This design not only reduces bias from self perception, but also allows for causal tests of whether prompting goal awareness can itself function as an intervention to enhance digital literacy and well-being.

Second, despite extensive efforts to maximize data coverage, missing data remain a limitation across multiple sources, including EMA surveys, smartphone screenshots, and TikTok data donation packages. In particular, objective behavioral analyses were restricted to TikTok because comparable donated data were unavailable for other platforms, and not all EMA-reported usage episodes could be linked to objective records. As a result, some episodes of social media

use, especially those involving non-TikTok platforms, are excluded from analyses on the objective behavioral signals for goal awareness and goal strength.

Third, our analyses, especially those involving objective TikTok behaviors, entailed multiple hypothesis tests aimed at identifying potential signals in high-dimensional data. Although this approach was appropriate for hypothesis generation and pattern discovery, many tests were not adjusted for multiple comparisons and should be interpreted cautiously. Future research will benefit from preregistered hypotheses, tighter model specifications, and confirmatory designs to rigorously evaluate the behavioral pathways suggested by the present findings. More specifically, we are extending our analyses of objective TikTok data by linking behavioral patterns with the content of viewed videos. Together with recent large-scale data donation efforts in research [213–215], these advances are expected to enable the development of theory-driven and interpretable models that move beyond purely exploratory data mining.

Fourth, the sample was purposively recruited via Prolific, an online research panel, with screening criteria targeting frequent TikTok users in the United States. As a result, the sample is not representative of the U.S. adult population. Specifically, participants were younger, more likely to be female, and more active on social media than the general population. As an opt-in convenience panel, the current study does not use a probability-based sampling, and sampling weights are not applicable given that the primary analyses examine within-person psychological processes rather than population-level estimates. Furthermore, participation in the TikTok data donation component introduces an additional layer of selection bias. Participants who consented to share their behavioral data may differ systematically from those who declined, for example in their comfort with data sharing, level of platform engagement, or attitudes toward social media use. This may limit the representativeness of the subsample used in the objective behavioral analyses. Generalizability of the findings to less frequent social media users, older adults, or populations outside the United States remains an open question. Future research would benefit from more diverse samples, including participants from non-Western contexts and lower social

media engagement groups, to establish the boundary conditions of the goal pursuit framework in digital environments.

3.5 Conclusion

This chapter examined social media use through the lens of momentary goal pursuit, combining a seven-day EMA with smartphone screenshots and TikTok data donation to capture both subjective experiences and objective behaviors in state-level social media use. Applying goal-pursuit theory to social media engagement, we found that higher levels of goal awareness and goal strength at the moment of use were consistently associated with happier, more meaningful social media experiences and greater happiness. Importantly, these associations were broadly robust across different types of social media use goals, suggesting that the awareness of goals matters regardless of whether users engage for entertainment, coping, information seeking, or habitual reasons.

At the same time, we found little evidence that higher goal awareness or strength translated into systematic differences in objective TikTok use patterns, such as usage duration, session frequency, video skipping, or active engagement. These objective patterns of TikTok use were not significantly associated with well-being outcomes, with the exception that people who watch a smaller proportion of each video before skipping tend to report higher happiness, high meaningfulness of social media experience, and higher level of goal strength. In general, these largely null findings highlight the complexity of real-world digital behavior and the limits of inferring psychological states from platform traces alone, while also underscoring the challenges of studying highly habitual and high-dimensional forms of engagement.

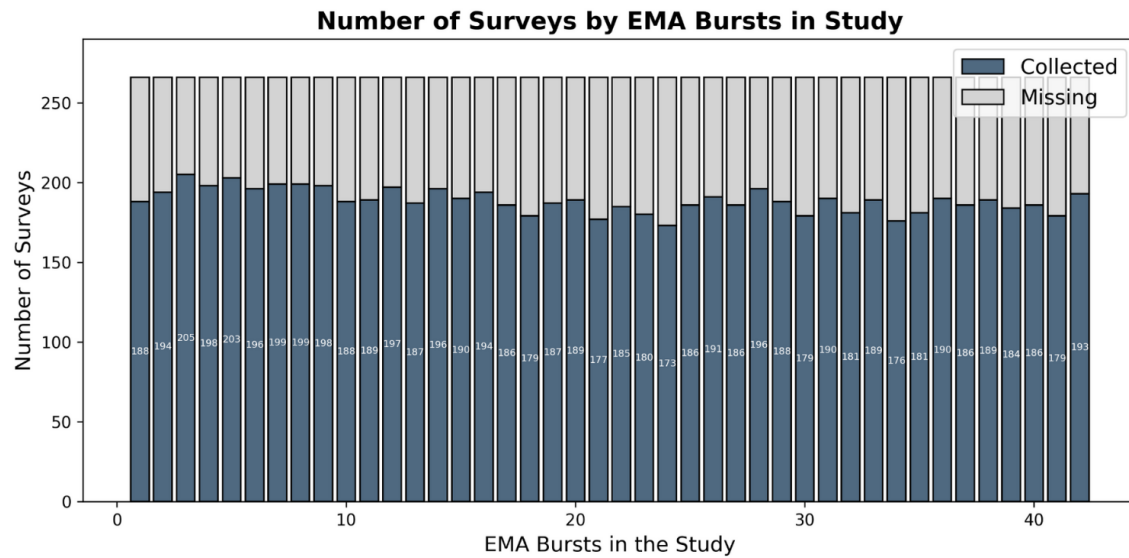
Taken together, these results suggest that supporting users' goal pursuit, particularly by enhancing goal awareness during use, may improve digital well-being. This perspective highlights promising opportunities for user-centered digital literacy interventions, while also raising deeper questions about how habitual use can be meaningfully integrated into goal-pursuit frameworks.

Beyond individual behavior, our findings underscore the responsibility of platforms to design systems that align with users' goals and support agency rather than merely maximizing engagement. At the same time, our empirical analyses illustrate both the promise and the difficulty of integrating subjective experience with objective digital trace data. Finally, the insights from this work extend beyond social media to emerging AI-driven systems, where aligning technology with human goals and well-being remains an urgent and unresolved challenge.

Supplementary materials

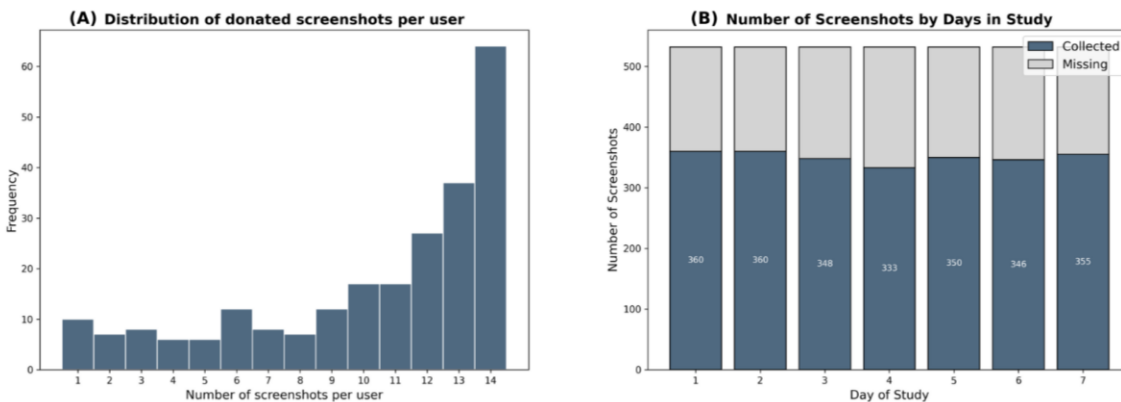
Figures

Figure A2.1. Number of surveys by EMA burst in the study



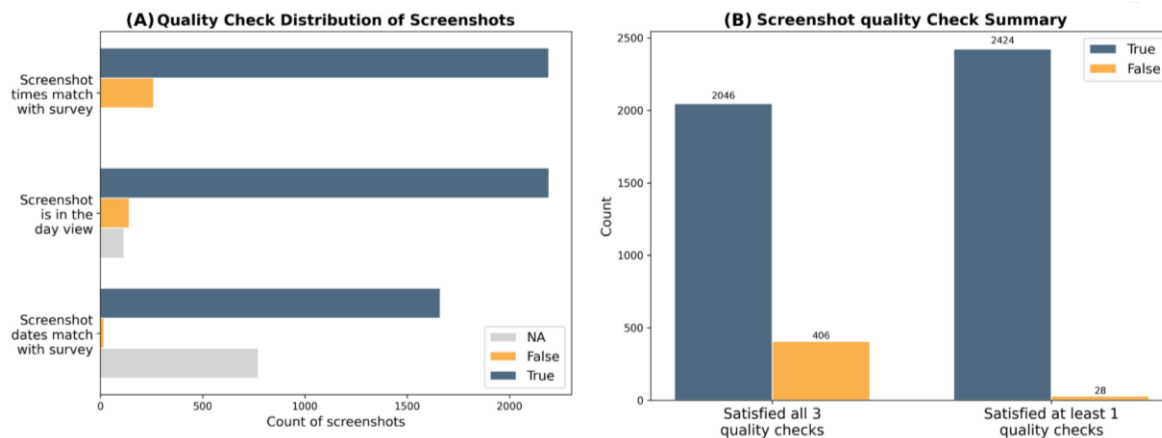
Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent.

Figure A2.2. The screenshot compliance rate



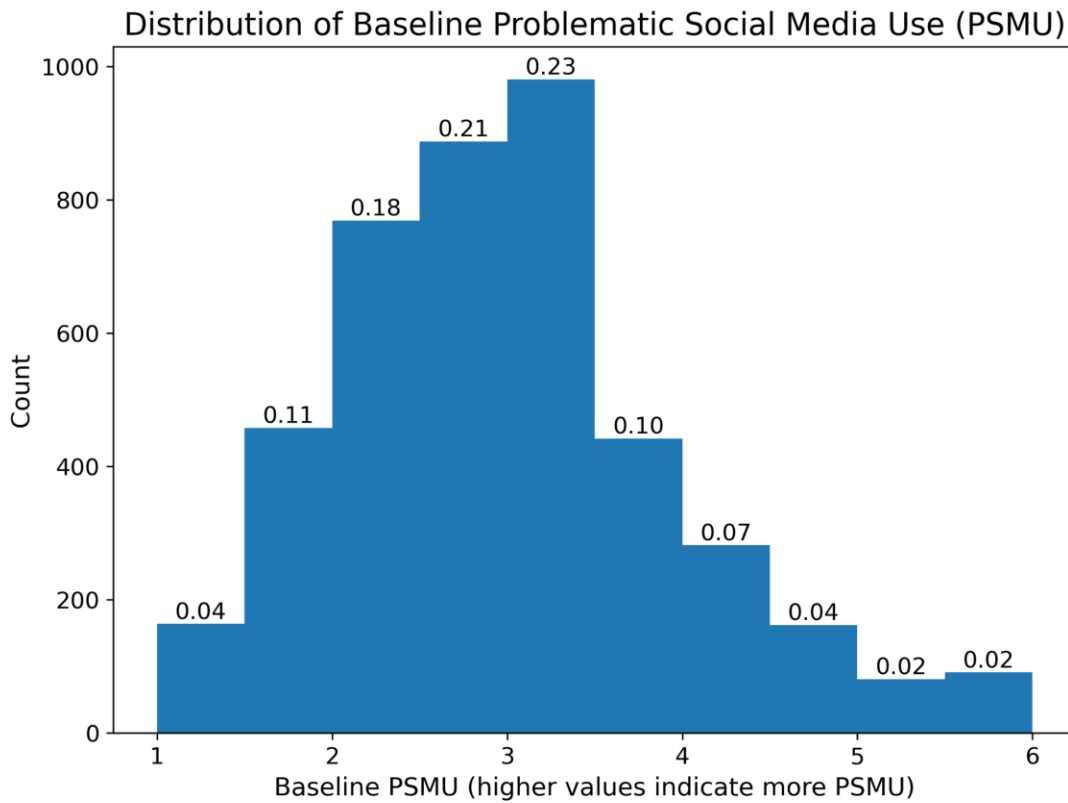
Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Objective social media use time data were collected via participant-submitted screenshots of smartphone screen-time summaries. Text of total screentime and app-specific screentime were extracted from screenshots.

Figure A2.3. The quality of screenshots



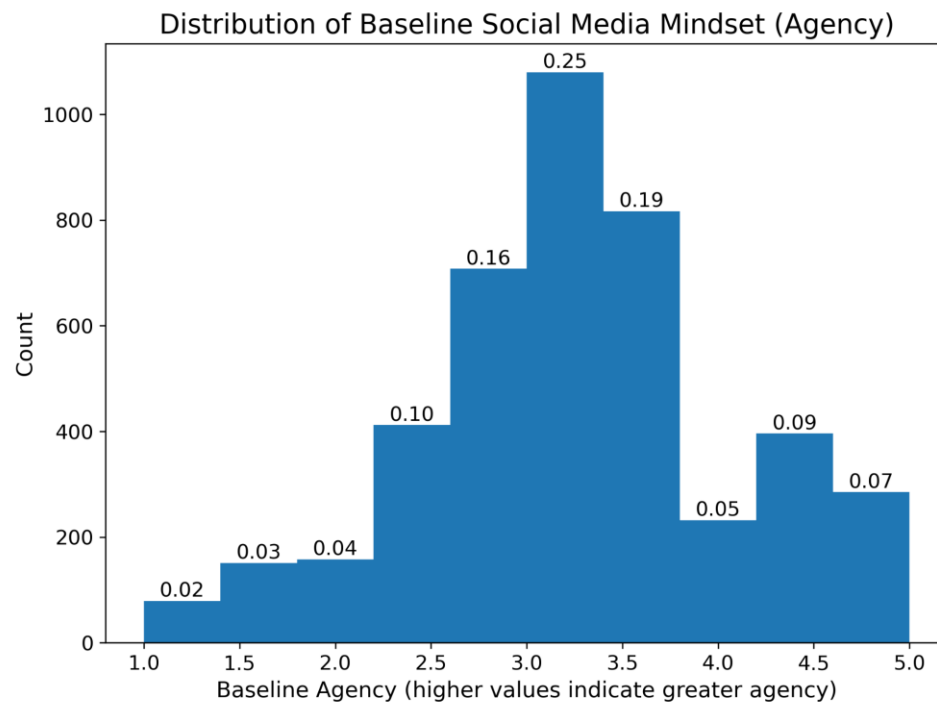
Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Objective social media use time data were collected via participant-submitted screenshots of smartphone screen-time summaries. Text of total screentime and app-specific screentime were extracted from screenshots.

Figure A2.4. Distribution of the baseline problematic social media use (PSMU)



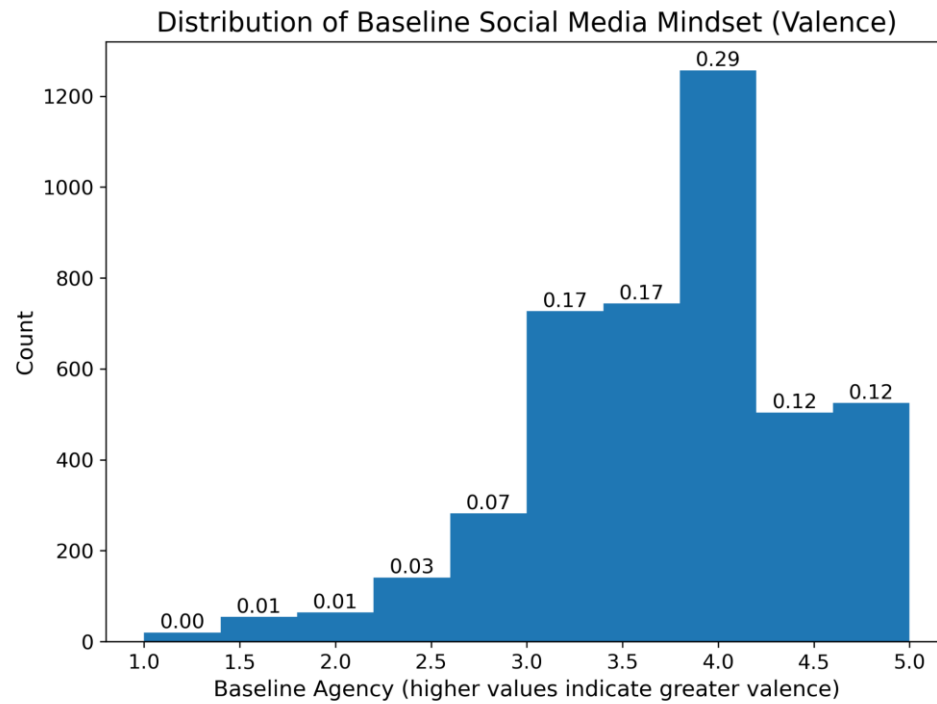
Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent.

Figure A2.5. Distribution of the baseline social media mindset (Agency)



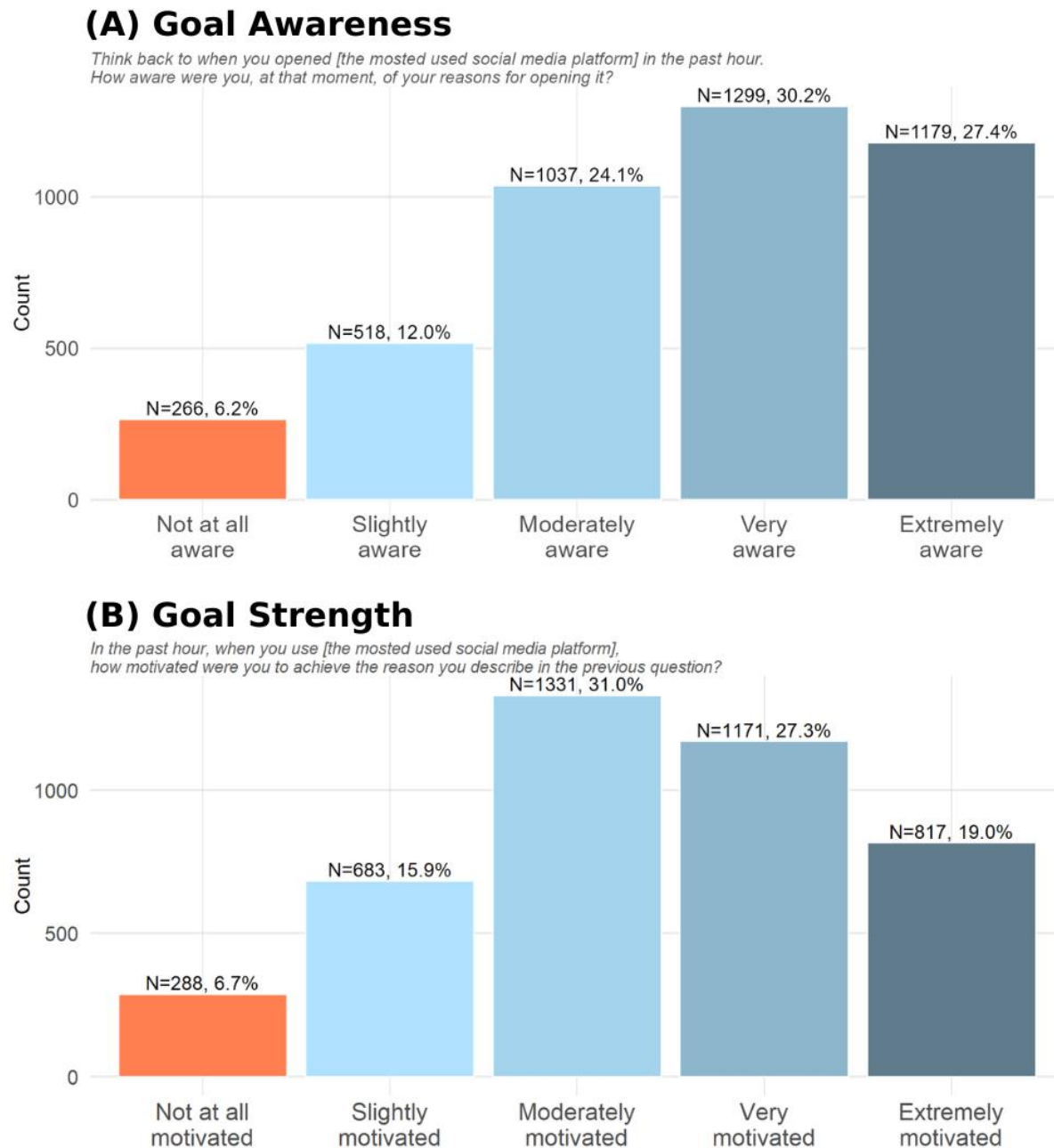
Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent.

Figure A2.6. Distribution of the baseline social media mindset (Valence)



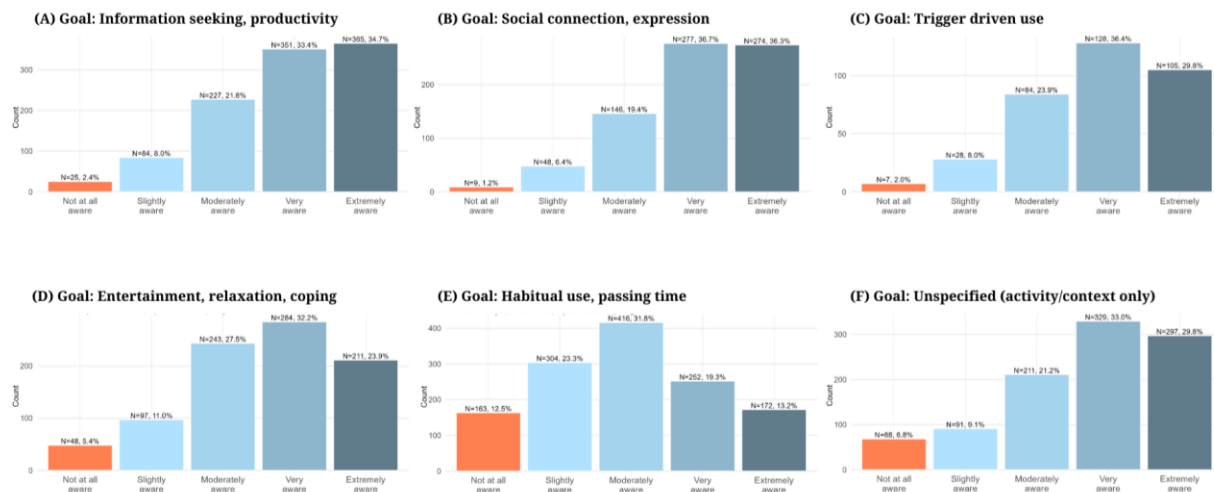
Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent.

Figure A2.7. The distribution of goal awareness and goal strength in the current study



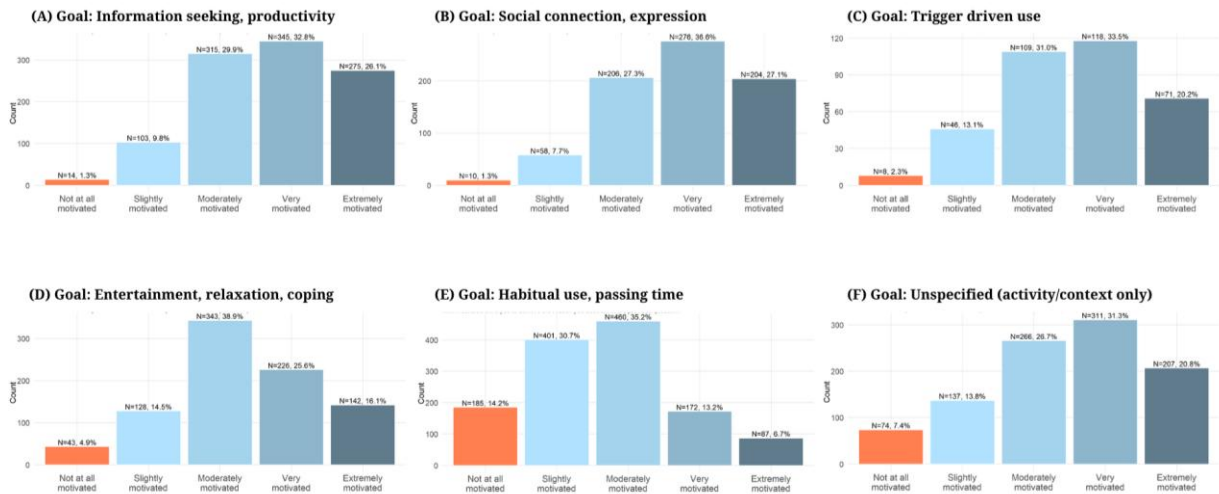
Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent.

Figure A2.8. Distribution of goal awareness across subgroups defined by self-reported social media use goals



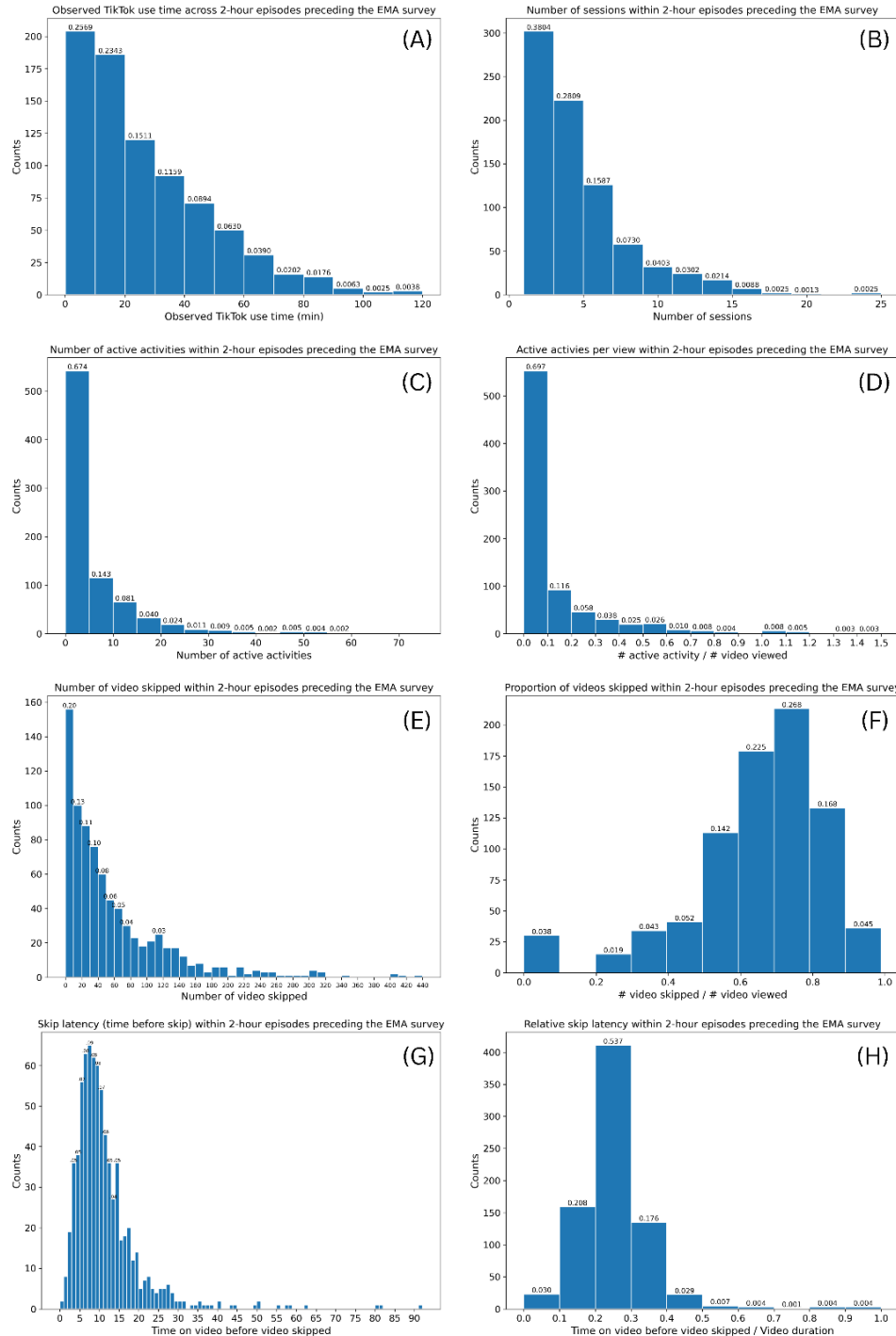
Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent.

Figure A2.9. Distribution of goal strength across subgroups defined by self-reported social media use goals



Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent.

Figure A2.10. Distribution of TikTok use behaviors, including observed time use (A), number of sessions (B), number of active activities (C), active activities per video view (D), number of video skips (E), proportion of videos skipped among all video views (F), time before video skipped (G), and time before video skip relative to total video duration (H), among 2-hour TikTok use episodes preceding EMA surveys

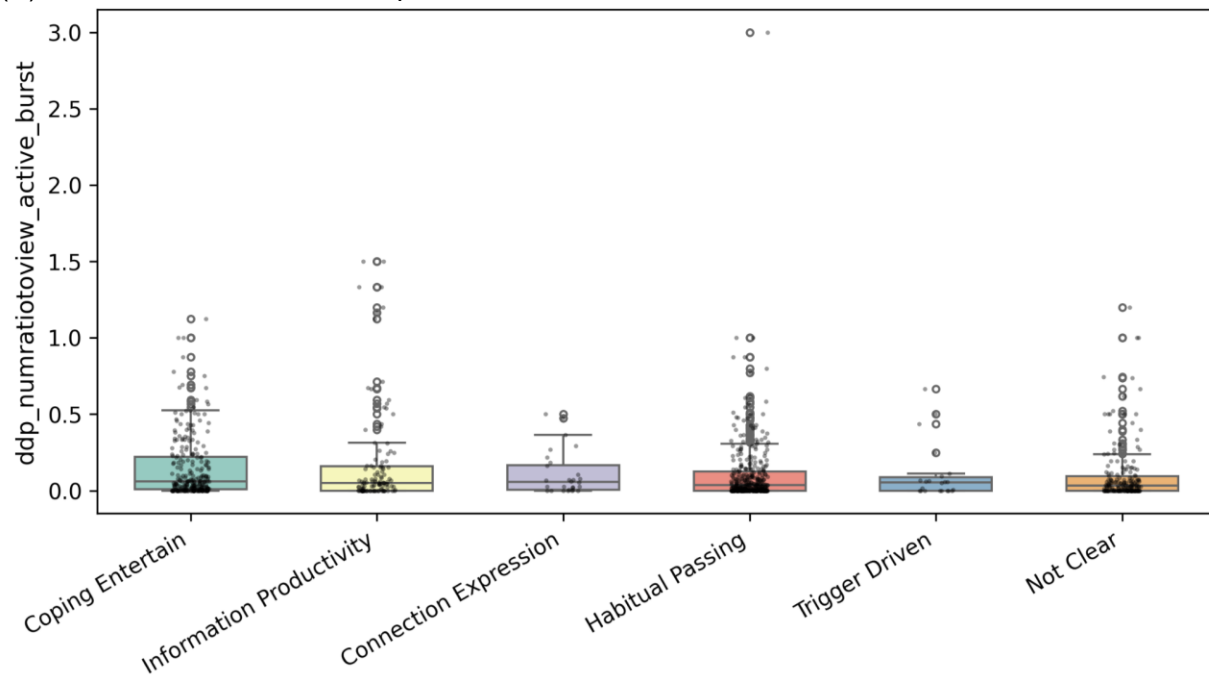


Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Donated TikTok data

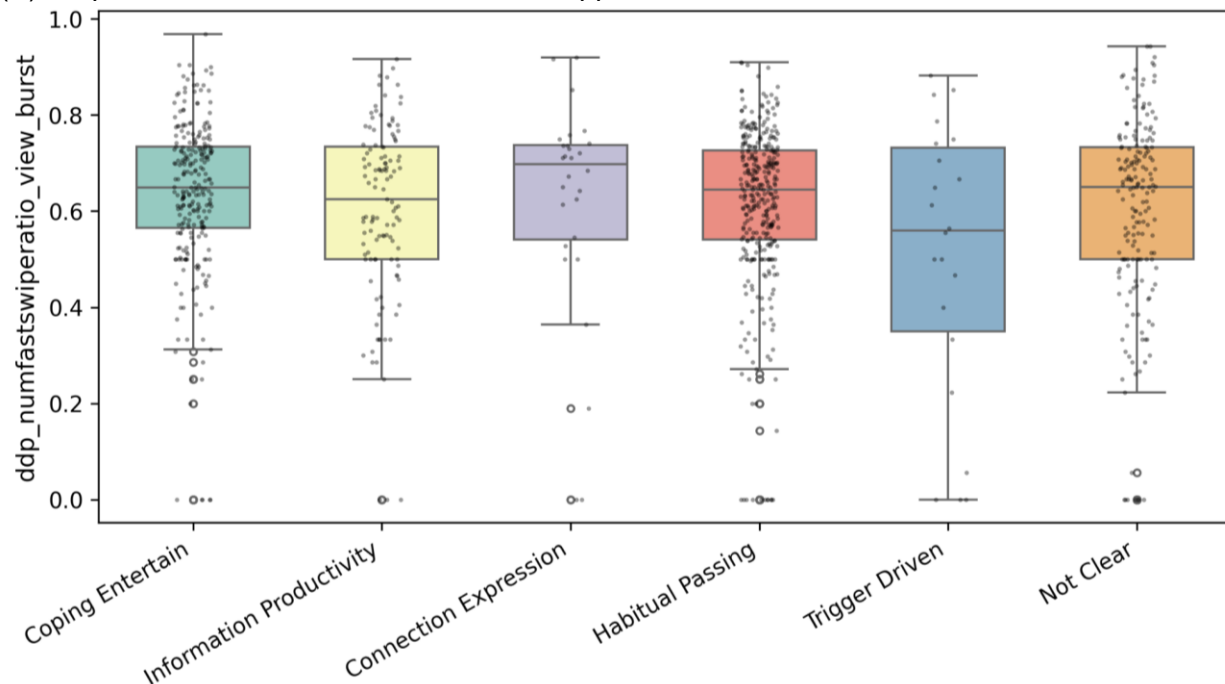
included time-stamped records of user activities and content interactions, which were linked to EMA reports and restricted to bursts in which TikTok was identified as the most-used app.

Figure A2.11. Active engagement and video skipping across self-reported tiktok use goals

(A) Number of active activities per video view

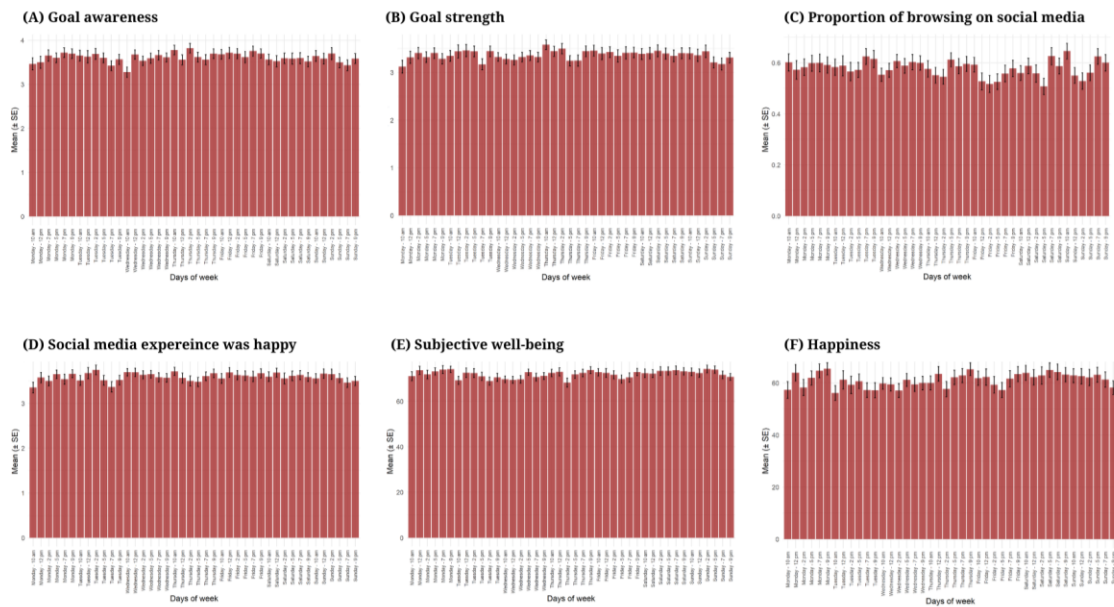


(B) Proportion of viewed videos that were skipped



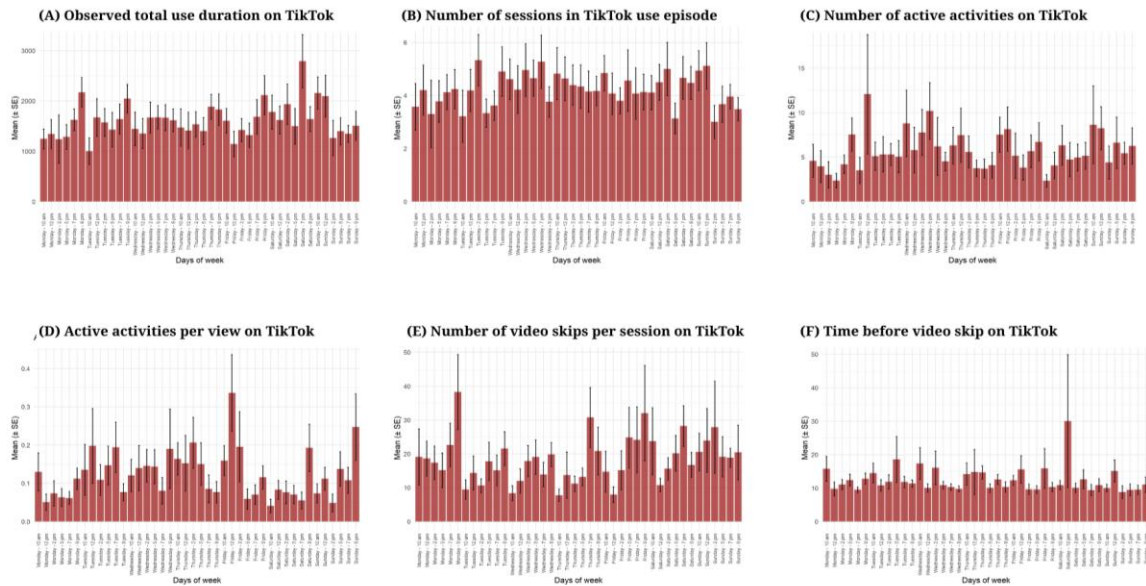
Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Donated TikTok data included time-stamped records of user activities and content interactions, which were linked to EMA reports and restricted to bursts in which TikTok was identified as the most-used app.

Figure A2.12. Temporal variation in self-reported social media experiences and activities



Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent.

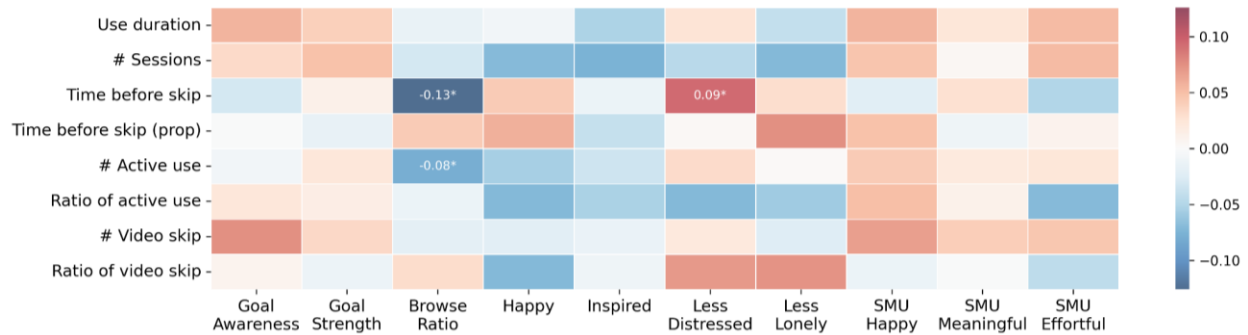
Figure A2.13. Temporal variation in objective tiktok use variables



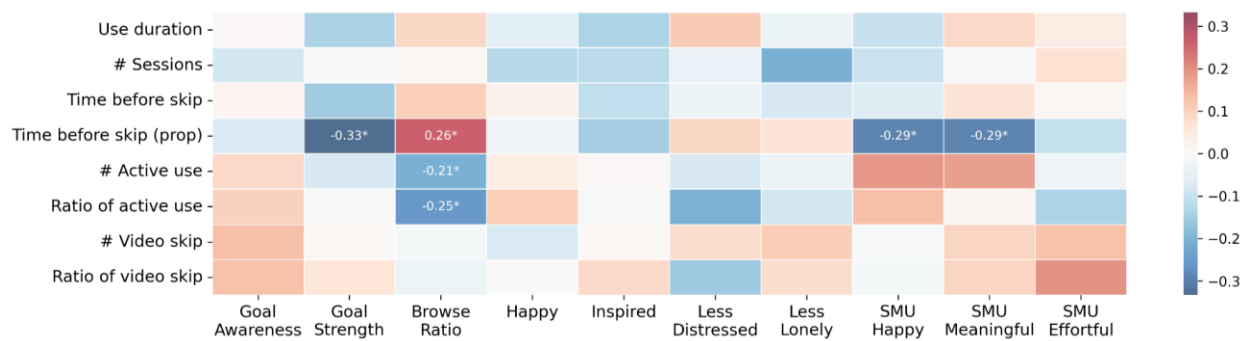
Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Donated TikTok data included time-stamped records of user activities and content interactions, which were linked to EMA reports and restricted to bursts in which TikTok was identified as the most-used app.

Figure A2.14. Within-person (A) and between-person (B) correlations between TikTok use behaviors and goal awareness, goal strength, social media use experience, and subjective well-being

(A) Within-person correlation



(B) Between-person correlation

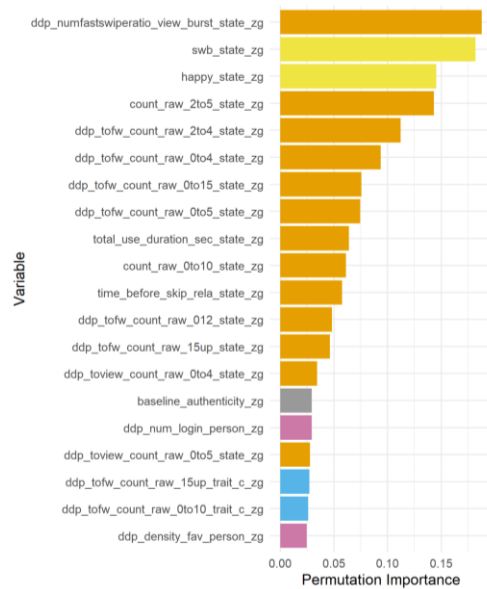


Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Donated TikTok data included time-stamped records of user activities and content interactions, which were linked to EMA reports and restricted to bursts in which TikTok was identified as the most-used app. Analyses reported are between-person and within-person Pearson correlations computed from person-level means and person-mean-centered deviations, respectively.

Figure A2.15. Important variables selected by random forest analysis

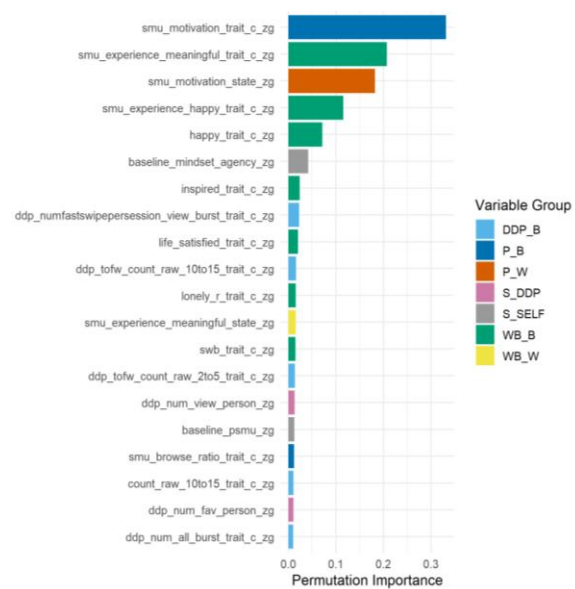
(A) Time before video skipped

Δ RMSE: 0.137



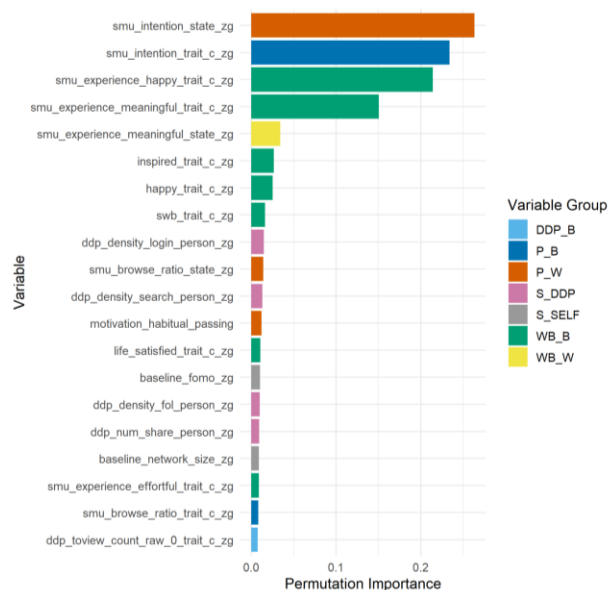
(B) Goal awareness of social media

Δ RMSE: 0.201



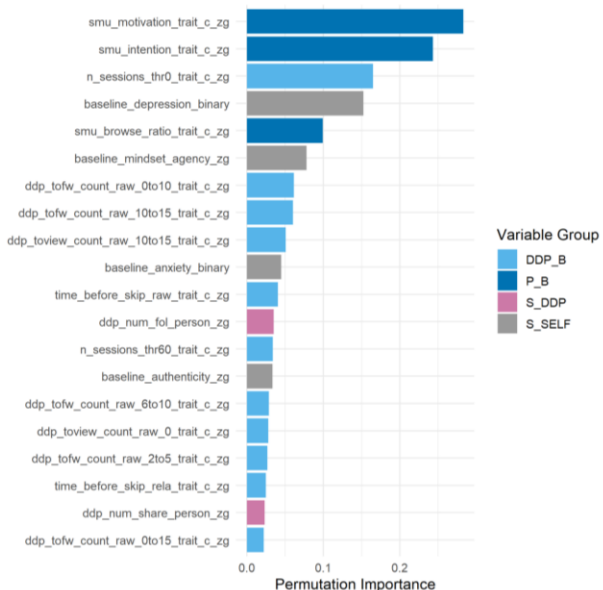
(C) Goal strength of social media use

Δ RMSE: 0.253



(D) Happiness (excluded well-being outcomes)

Δ RMSE: 0.085

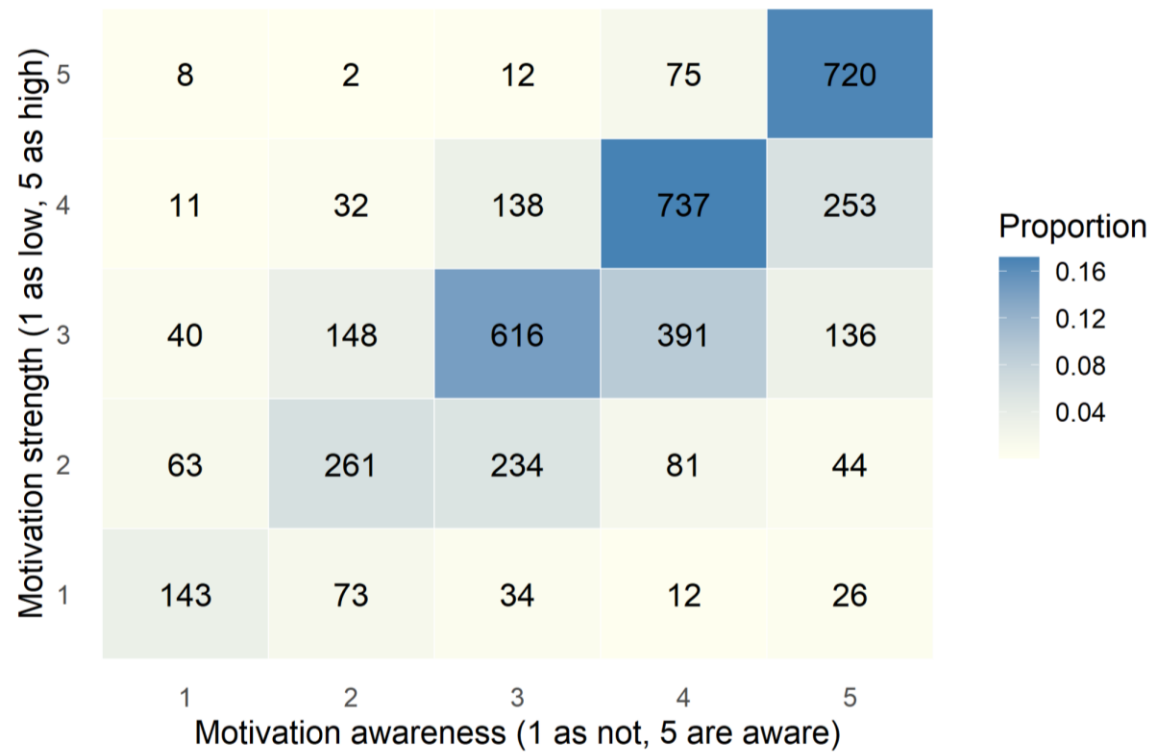


Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Donated TikTok data included time-stamped records of user activities and content interactions, which were linked to EMA reports and restricted to bursts in which TikTok was identified as the most-used app. Predictor importance was assessed using random forest models with Hybrid/Mundlak feature decomposition to capture nonlinear, high-dimensional associations while disentangling within- and between-person effects.

- p_w: Within-person EMA survey measures, including momentary independent variables (goal awareness, goal strength) and time-varying covariates related to social media use and goals.

- p_b: Between-person (person-mean) versions of EMA survey variables, capturing stable individual differences in goals and social media use reported during the EMA period.
- s_self: Stable person-level characteristics, including demographics, baseline psychological traits (e.g., anxiety, depression), social media use mindset, need for cognition, problematic social media use, total screen time, and dominant platform use.
- s_ddp: Long-term person-level TikTok behavioral indicators derived from the full data donation history (e.g., total activity count, average daily activity since platform adoption).
- ddp_b: Person-level TikTok activity aggregated over the 14-day EMA window.
- ddp_w: Within-person fluctuations in TikTok activity during the EMA period.
- wb_b: Between-person measures of subjective experience and well-being (person means across EMA).
- wb_w: Within-person deviations in subjective experience and well-being during EMA.
- todw: Temporal context variables, including day of week and time of day.

Figure A2.16. Heatmap of the distribution of goal awareness and goal strength in social media use



Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent.

Tables

Table A2.1. Performance of text classifier for social media use motivation

	Example	N (%)	Cohen κ	Precision	Recall	F1
Coping	To distract myself.	247 (0.06)	0.77	0.88	0.71	0.79
Entertainment	Scrolling for entertainment.	717 (0.17)	0.75	0.66	0.97	0.78
Habitual use	I used it because it is an addictive force of habit.	399 (0.09)	0.64	0.67	0.67	0.67
Info seeking	Looking for a show to watch.	995 (0.23)	0.76	0.80	0.81	0.81
Passing time	I want to kill time.	1059 (0.25)	0.90	0.92	0.92	0.92
Trigger driven	got a notification	113 (0.03)	0.72	0.79	0.69	0.73
Productivity	I used it to create a WFH video.	142 (0.03)	0.87	0.78	0.99	0.88
Self-expression	I use it to post and share my thoughts.	648 (0.15)	0.70	0.55	0.99	0.71
Social connection	Contacting friends via DMs/messages.	352 (0.08)	0.67	0.74	0.70	0.72
Activity / context only	I wanted to watch some videos.	996 (0.23)	0.69	0.70	0.79	0.74

Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Open-ended responses in each EMA survey describing reasons for social media use were classified into 10 intention categories using an LLM-assisted coding approach (GPT-4.0) guided by a literature-based codebook, with reliability assessed via comparison to human coding.

Table A2.2. Demographic characteristics of the study sample (N=4,318)

Variable	Statistics
Age	38.1 (11.9)
Sex, female (%)	61.5
Ethnicity, White (%)	69.1
Ethnicity, Black (%)	17.1
Ethnicity, Asian (%)	6.2
Ethnicity, Others (%)	7.4
Is student (%)	17.9
Unemployed (%)	4.1
Country of birth, Non US (%)	5.3

Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent.

Table A2.3. Descriptive statistics for key variables measured in the EMA study

	N (EMA)	N (person)	Scale (Min--Max)	Mean	SD	P25	P50	P75	Between- Person SD	Within- Person SD	ICC
Goal pursuit variables											
Goal awareness	4299	263	1--5	3.606	1.183	3	4	5	0.881	0.854	0.516
Goal strength	4290	263	1--5	3.36	1.155	3	3	4	0.829	0.846	0.490
SMU experience in the past hour was:											
Happy	4286	263	1--5	3.591	1.042	3	4	4	0.846	0.694	0.598
Meaningful	4286	263	1--5	3.411	1.195	3	4	4	0.934	0.8	0.577
Effortful	4286	263	1--5	2.409	1.308	1	2	3	1.076	0.784	0.653
Subjective Well-being											
Happy	4292	263	0--100	61.129	28.58	44	64	84	25.233	14.764	0.745
Angry (r*)	4292	263	0--100	4.728	13.693	0	0	0	13.358	10.592	0.614
Inspired	4290	263	0--100	36.535	33.863	0	29	64	28.866	16.967	0.743
Lonely (r*)	4288	263	0--100	13.54	25.442	0	0	14	22.268	13.278	0.738
Distressed (r*)	4288	263	0--100	9.927	19.274	0	0	11	16.1	13.082	0.602
Life satisfaction	4288	263	0--100	61.66	28.559	44	65	85	26.467	13.298	0.798
Level of social media use activity engagement in the past hour:											
Posting	4305	263	0--60	3.931	10.462	0	0	0	9.404	7.5	0.611
Browsing	4304	263	0--60	20.451	16.426	8	16	30	12.563	12.596	0.499
Searching	4303	263	0--60	7.155	12.277	0	1	9	11.045	7.596	0.679
Messaging	4302	263	0--60	4.525	10.872	0	0	3	8.938	7.37	0.595
Commenting	4101	262	0--60	7.432	12.032	0	2	10	10.759	7.099	0.697

*r: Reverse coded

Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent.

Table A2.4. Descriptive statistics for screentime from the user shared screenshots in the current study

	N (EMA)	N (person)	Scale (Min--Max)	Mean	SD	P25	P50	P75	Between- Person SD	Within- Person SD	ICC
Total screen time	1270	235	0--1433	430.615	274.163	268	385.5	538	242.57	204.103	0.585
By app categories											
Social	1270	235	0--1363	166.196	158.59	56	136.5	233	129.109	108.697	0.585
Productivity	1270	235	0--1237	56.782	114.581	0	9	68	89.279	85.451	0.522
Entertainment	1270	235	0--1390	45.251	122.02	0	0	22	89.259	83.767	0.532
Games	1270	235	0--1378	33.5	100.514	0	0	0	118.209	70.592	0.737
Utilities	1270	235	0--954	30.343	86.616	0	0	8	66.881	61.654	0.541
Information	1270	235	0--948	21.025	69.813	0	0	0	43.332	58.569	0.354
Shop	1270	235	0--507	9.333	38.062	0	0	0	23.157	35.431	0.299
Others	1270	235	0--849	24.131	75.03	0	0	0	53.501	65.705	0.399
By apps											
TikTok	1270	235	0--1019	52.092	105.708	0	0	65.75	89.097	69.636	0.621
Browser	1270	235	0--1308	71.427	126.103	0	16	94	96.158	87.748	0.546
Game apps	1270	235	0--1358	30.174	98.917	0	0	0	126.506	64.607	0.793
Message apps	1270	235	0--1384	29.322	85.611	0	0	19	71.354	59.72	0.588
YouTube	1270	235	0--1429	25.357	95.433	0	0	0	70.421	59.767	0.581
Facebook	1270	235	0--750	25.209	57.474	0	0	29	47.272	41.422	0.566
Instagram	1270	235	0--897	16.62	52.229	0	0	0	34.573	39.811	0.43
Mail apps	1270	235	0--899	11.836	63.311	0	0	0	46.589	48.893	0.476
ChatGPT	1270	235	0--836	5.964	48.272	0	0	0	36.041	29.508	0.599

Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Objective social media use time was derived from participant-submitted smartphone screen-time screenshots processed via an established text-recognition pipeline, with quality checks applied to ensure temporal alignment and data validity.

Table A2.5. Descriptive statistics of activities on TikTok, extracted from user donated TikTok data

	N (EMA)	N (person)	Scale (Min-- Max)	Mean	SD	P25	P50	P75	Between- Person SD	Within- Person SD	ICC
Total use duration (sec)	794	124	7--8,455	1626.441	1336.802	575.75	1264.52	379.75	798.443	1126.69	0.334
Number of sessions*	794	124	1--25	4.237	3.389	2	3	6	2.016	2.565	0.382
Basic TikTok activities (frequency)											
All activities	807	124	1--696	88.691	86.816	28	65	122.5	47.629	69.669	0.319
View	807	124	0--692	78.802	82.184	22.5	53	110.5	44.979	64.637	0.326
Like	633	101	0--87	4.468	9.123	0	1	5	6.367	6.764	0.47
Login	807	124	0--38	4.2	4.243	1	3	6	2.555	2.974	0.425
Search	804	123	0--20	1.218	2.095	0	0	2	1.343	1.792	0.36
Share	774	114	0--9	0.309	1.016	0	0	0	0.8	0.8	0.5
Comment	579	90	0--15	0.247	1.276	0	0	0	0.701	0.851	0.404
Post	291	45	0--1	0.007	0.083	0	0	0	0.05	0.1	0.203
Follow	767	115	0--5	0.156	0.53	0	0	0	0.452	0.508	0.441
Favorite	740	109	0--15	0.38	1.286	0	0	0	1.131	1.088	0.519
Active activities (like, follow, favorite, search, comment, share, and post)											
Number of active events	807	124	0--110	5.69	9.748	0	2	7	6.9	7.473	0.46
Active activities per view	794	124	0--3	0.126	0.234	0	0.038	0.143	0.259	0.266	0.486
Active activities per minute of use	794	124	0--7.83	0.29	0.603	0	0.109	0.348	0.625	0.784	0.389
Video skip behaviors											
Number of video skips	794	124	0--613	59.252	68.815	14	36	79.75	37.816	52.705	0.34
Video skips per view	794	124	0--0.989	0.646	0.197	0.568	0.687	0.776	0.16	0.189	0.419
Video skips per minute of use	794	124	0--14	2.141	1.753	0.951	1.812	2.734	1.307	1.31	0.499
Skip latency (time to skip, in sec)	765	121	0--368	12.457	17.389	6.5	9.452	14.043	7.77	16.735	0.177
Skip latency (time to skip/video duration, in prop)	765	121	0--0.957	0.254	0.101	0.204	0.247	0.293	0.065	0.096	0.315

Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Donated TikTok data included time-stamped records of user activities and content interactions, which were linked to EMA reports and restricted to bursts in which TikTok was identified as the most-used app.

Table A2.6. Associations of goal awareness and goal strength with subjective well-being and social media use experience, by subgroups of momentary goals

	Coping Entertain	Information Productivity	Connection Expression	Habit Pass time	Response to triggers	Goal is unclear
Outcome: Happy						
Goal awareness (w)	0.07 (0.02)*	0.03 (0.02)	-0.01 (0.03)	0.04 (0.02)*	-0.05 (0.04)	0.08 (0.03)*
Goal awareness (b)	0.18 (0.09)*	0.16 (0.10)	0.04 (0.12)	0.21 (0.09)*	0.05 (0.12)	0.16 (0.10)
Goal strength (w)	0.04 (0.03)	0.07 (0.02)*	0.06 (0.03)*	0.03 (0.02)	0.08 (0.03)*	0.08 (0.03)*
Goal strength (b)	0.28 (0.10)*	0.35 (0.10)*	0.39 (0.12)*	0.35 (0.09)*	0.44 (0.12)*	0.33 (0.10)*
Outcome: Inspired						
Goal awareness (w)	0.07 (0.03)*	0.02 (0.02)	-0.04 (0.03)	0.01 (0.02)		0.04 (0.02)
Goal awareness (b)	0.08 (0.09)	0.08 (0.09)	-0.00 (0.11)	0.01 (0.08)		-0.00 (0.10)
Goal strength (w)	0.05 (0.02)*	0.08 (0.03)*	0.14 (0.04)*	0.05 (0.02)*		0.07 (0.03)*
Goal strength (b)	0.37 (0.09)*	0.31 (0.09)*	0.37 (0.11)*	0.39 (0.09)*		0.35 (0.10)*
Outcome: Less distressed						
Goal awareness (w)	0.03 (0.03)				-0.06 (0.06)	0.02 (0.02)
Goal awareness (b)	0.06 (0.10)				0.19 (0.13)	0.16 (0.11)
Goal strength (w)	-0.00 (0.04)				0.10 (0.08)	0.03 (0.03)
Goal strength (b)	-0.05 (0.10)				0.02 (0.13)	-0.18 (0.11)
Outcome: Less lonely						
Goal awareness (w)	0.05 (0.02)*	0.03 (0.02)	-0.05 (0.03)	0.06 (0.02)*	-0.05 (0.04)	-0.01 (0.02)
Goal awareness (b)	-0.03 (0.10)	0.08 (0.11)	-0.12 (0.12)	-0.06 (0.10)	0.03 (0.13)	0.05 (0.11)
Goal strength (w)	-0.02 (0.02)	-0.00 (0.02)	0.04 (0.03)	-0.00 (0.02)	0.03 (0.04)	0.01 (0.02)
Goal strength (b)	0.19 (0.11)	0.11 (0.10)	0.24 (0.12)	0.22 (0.11)*	0.23 (0.13)	0.11 (0.12)
SMU experience is happy						
Goal awareness (w)	0.10 (0.03)*	0.02 (0.03)		0.07 (0.03)*		0.06 (0.03)
Goal awareness (b)	0.01 (0.07)	0.01 (0.08)		0.02 (0.07)		0.05 (0.07)
Goal strength (w)	0.16 (0.04)*	0.17 (0.03)*		0.13 (0.03)*		0.21 (0.03)*
Goal strength (b)	0.51 (0.07)*	0.47 (0.08)*		0.49 (0.08)*		0.41 (0.07)*
SMU experience is meaningful						
Goal awareness (w)	0.08 (0.03)*	0.06 (0.03)	0.03 (0.03)	0.07 (0.02)*		0.03 (0.03)
Goal awareness (b)	0.10 (0.07)	0.04 (0.07)	-0.09 (0.08)	0.08 (0.07)		0.11 (0.07)
Goal strength (w)	0.17 (0.03)*	0.20 (0.03)*	0.26 (0.04)*	0.13 (0.02)*		0.20 (0.04)*
Goal strength (b)	0.45 (0.07)*	0.42 (0.07)*	0.48 (0.08)*	0.40 (0.08)*		0.39 (0.07)*
SMU experience is effortful						
Goal awareness (w)	0.04 (0.03)	0.06 (0.03)*		0.01 (0.02)		0.01 (0.02)
Goal awareness (b)	-0.09 (0.09)	-0.04 (0.10)		-0.03 (0.08)		-0.08 (0.11)
Goal strength (w)	-0.02 (0.03)	0.03 (0.03)		0.03 (0.02)		0.05 (0.03)
Goal strength (b)	0.25 (0.10)*	0.22 (0.10)*		0.15 (0.09)		0.30 (0.11)*

Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. Reported results are derived from separate three-level linear mixed-effects models within subsets defined by social media use goal type, using identical model specifications to those in the main analyses in H1. * indicates P value < 0.05. Results from models that fail to converge in estimation were not presented.

Table A2.7. Associations of social media goal awareness and strength with happiness, moderated by group subgroups

	Coping & Entertainment	Information & Productivity	Connection & Expression	Habitual use & Passing time	Response to triggers	Goal unclear
Goal awareness (within)	0.03 (0.01)*	0.04 (0.01)*	0.05 (0.01)*	0.04 (0.01)*	0.04 (0.01)*	0.03 (0.01)*
Goal awareness (between)	0.11 (0.08)	0.10 (0.08)	0.11 (0.08)	0.12 (0.08)	0.11 (0.08)	0.11 (0.08)
Goal strength (within)	0.06 (0.01)*	0.06 (0.01)*	0.05 (0.01)*	0.07 (0.01)*	0.06 (0.01)*	0.06 (0.01)*
Goal strength (between)	0.34 (0.08)*	0.35 (0.08)*	0.36 (0.08)*	0.35 (0.08)*	0.35 (0.08)*	0.36 (0.08)*
Goal subgroup	-0.04 (0.02)	-0.02 (0.02)	0.08 (0.03)*	0.00 (0.02)	0.08 (0.03)*	-0.03 (0.02)
Awareness (within) x goal subgroup	0.03 (0.02)	-0.01 (0.02)	-0.07 (0.03)*	0.00 (0.02)	-0.05 (0.04)	0.03 (0.02)
Awareness (between) x goal subgroup	0.01 (0.03)	0.03 (0.03)	0.03 (0.04)	-0.01 (0.03)	-0.03 (0.05)	0.01 (0.04)
Strength (within) x goal subgroup	0.00 (0.02)	0.01 (0.02)	0.04 (0.03)	-0.03 (0.02)	0.03 (0.03)	0.02 (0.02)
Strength (between) x goal subgroup	0.03 (0.04)	0.01 (0.03)	-0.05 (0.04)	0.02 (0.03)	-0.02 (0.05)	-0.02 (0.04)

Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. Reported results are derived from separate three-level linear mixed-effects models within subsets defined by social media use goal type, using identical model specifications to those in the main analyses in H1. * indicates P value < 0.05. Estimates are from three-level linear mixed-effects models with interaction terms testing whether social media use goal type moderates within-person and between-person associations between goal awareness, goal strength, and outcomes.

Table A2.8. Mediation analysis of goal awareness and goal strength on happy social media experience via objective tiktok use

X = Goal awareness		X = Goal strength	
term	Estimate (SE)	term	Estimate (SE)
<i>m = Number of active activities</i>			
dm:time_c	0.02 (0.02)	dm:time_c	0.02 (0.02)
time_c:dy	-0.01 (0.0)*	time_c:dy	-0.01 (0.0)*
dm:X	-0.02 (0.34)	dm:X	0.19 (0.33)
dy:X	0.21 (0.04)*	dy:X	0.25 (0.04)*
dy:M	0.0 (0.0)	dy:M	0.0 (0.0)
<i>m = Ratio of view skipped</i>			
dm:time_c	0.0 (0.0)	dm:time_c	0.0 (0.0)
time_c:dy	-0.01 (0.0)*	time_c:dy	-0.01 (0.0)*
dm:X	0.0 (0.01)	dm:X	-0.0 (0.01)
dy:X	0.22 (0.04)*	dy:X	0.25 (0.04)*
dy:M	0.01 (0.15)	dy:M	0.04 (0.16)
<i>m = Time before video skipped relative to video duration</i>			
dm:time_c	0.0 (0.0)	dm:time_c	0.0 (0.0)
time_c:dy	-0.01 (0.0)*	time_c:dy	-0.01 (0.0)*
dm:X	-0.0 (0.01)	dm:X	-0.0 (0.01)
dy:X	0.2 (0.04)*	dy:X	0.24 (0.03)*
dy:M	0.45 (0.42)	dy:M	0.43 (0.42)

Note: Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Donated TikTok data included time-stamped records of user activities and content interactions, which were linked to EMA reports and restricted to bursts in which TikTok was identified as the most-used app. Indirect effects were tested using multilevel mediation analyses estimated with linear mixed-effects models, examining whether objective TikTok use variables mediated within-person associations between goal-related predictors and happiness, with time included as a covariate. Specifically:

- dm:time_c: The coefficient for dm:time_c represents the within-person time trend in objective TikTok use, indicating whether individuals' TikTok use systematically increases or decreases over the study period.
- time_c:dy: The coefficient for time_c:dy represents the within-person time trend in social media experience being happy, indicating whether reported happiness changes over time independent of predictors.
- dm:X: The coefficient for dm:X represents the a path, indicating the extent to which goal awareness (or goal strength) predicts within-person variation in objective TikTok use.
- dy:M: The coefficient for dy:M represents the b path, indicating whether within-person variation in objective TikTok use is associated with happy social media experience.
- dy:X: The coefficient for dy:X represents the direct effect (c' path) of goal awareness (or goal strength) on happy social media experience, adjusting for objective TikTok use.

Table A2.9. Results from linear regression models predicting change in agency and valence from baseline characteristics

	Change in Agency	Change in Valence
Intercept	-0.00 (0.01)*	-0.00 (0.01)
Age	0.03 (0.02)	0.07 (0.01)*
Female	-0.05 (0.01)*	-0.01 (0.01)
Anxiety	-0.09 (0.02)*	-0.12 (0.02)*
Depression	-0.16 (0.02)*	0.02 (0.02)
FOMO	-0.03 (0.02)	-0.04 (0.02)*
Baseline agency mindset	-0.53 (0.02)*	-0.00 (0.02)
Baseline valence mindset	0.04 (0.02)*	-0.57 (0.02)*
Problematic social media use	-0.03 (0.02)	0.05 (0.02)*
Social network size	0.02 (0.02)	0.12 (0.01)*
Need for cognition	0.19 (0.01)*	0.13 (0.01)*
Most used app: browser	0.10 (0.02)*	0.18 (0.02)*
Most used app: game	0.13 (0.02)*	0.12 (0.02)*
Most used app: messaging apps	0.07 (0.02)*	0.16 (0.01)*
Most used app: TikTok	0.08 (0.02)*	0.21 (0.02)*
Screentime on smartphone	0.09 (0.01)*	0.18 (0.01)*
Adjusted R square	0.288	0.386

Note: Note: Data come from a 7-day Ecological Momentary Assessment (EMA) study conducted in July 2025. Participants (N = 266; mean age = 37.9 years; 61.3% female) were English-speaking U.S. adults recruited via Prolific who had used TikTok at least once in the past month. All participants provided informed consent. Exploratory analyses examined individual-level changes in agency and valence mindsets from baseline to post-EMA and used linear regression models to identify baseline predictors of these changes.

Codebook

Section 1: LLM-assisted content analysis

codebook_type_of_use = {

"M1": {

"label": "Information Seeking",

"description": "Using social media to actively search for, look for, or find information, content, or updates. This includes seeking knowledge, tutorials, recommendations, shopping information, or inspiration, as well as checking for what is new or happening (e.g., events, news, updates from creators). The defining feature is that the user is deliberately seeking something, rather than casually consuming content for fun or habit.",

"examples": [

"Looking for a show to watch.",

"To see what happened at Tomorrowland (main stage burned down).",

"I was looking at videos and posts of the band I saw from other people that was there.",

"Needed a tutorial.",

"Shopping for clothes.",

"I was looking for inspiration but found nothing.",

"To see what's new going on.",

"To check if someone posted.",

"Wanted to see a few of my faves and if they posted.",

"To keep up with accounts I follow.",

"To get updates."

]

},

"M2":{

"label": "Social Connection",

"description": "Using social media to connect with one's own social network, such as friends or family.",

"examples": [

"Contacting friends via DMs/messages.",

"I have went on to keep up with the news about Trump and my friends.",

"To keep up with friends and family and stay in the know."

]

},

"M3": {

"label": "Self-Expression",

"description": "Using social media to actively express oneself, share one's own perspective, or engage in interactions that reflect one's thoughts. This includes creating posts, uploading content, commenting, interacting and participating in visible ways on the platform. It does not include passively viewing others' content, seeing updates from others, or replying to individual messages.",

"examples": [

```

    "I use it to post and share my thoughts.",
    "To express myself on IG.",
    "I wanted to post about the passing of Ozzy Osborne.",
    "I commented on a video",
    "I made a morning post",
  ]
},

"M4": {
  "label": "Entertainment / Relaxation",
  "description": "Using social media explicitly for fun, amusement, casual leisure, or lighthearted relaxation. This includes engaging with humorous, playful, relaxing, or entertaining content, or using the platform as background company. To be classified as Entertainment / Relaxation, a response must clearly indicate enjoyment, fun, relaxation, or amusement. Do not assign to this category if the response only (1) describes an activity without expressing enjoyment (e.g., 'I am watching videos' → should be M10; 'to watch videos and news' → should be M10 and M1), (2) refers to coping with negative states such as stress, fatigue, or illness (see M6), or (3) describes news, tutorials, or other information-seeking purposes (see M1).",
  "examples": [
    "Scrolling for entertainment.",
    "Basically I use it for fun.",
    "To enjoy watching funny videos.",
    'I use it to relax after a whole day busy work.',
    "Watching dance challenges on TikTok for fun.",
    "It is very interesting.",
    "Checking memes that make me laugh.",
    'Needed some background noise while I worked on something'
  ]
},

"M5": {
  "label": "Passing Time",
  "description": "Using social media to fill idle moments or alleviate boredom.",
  "examples": [
    "I want to kill time.",
    "To cure boredom.",
    "I was commuting",
    "I was in the restroom",
    'Passing time',
    'listening to podcasts while driving',
    'To look at videos and pass the time',
    'I was finishing lunch and looking for something to read.'
  ]
},

"M6": {

```

```

"label": "Coping",
"description": "Using social media to distract, comfort, or regulate oneself in response to negative, tiring, or stressful states. This includes using social media to take a break, to unwind, to relieve boredom, to deal with fatigue, or to manage negative emotions or difficult circumstances.",
"examples": [
    "To distract myself.",
    "I have used it because it is an addictive force of habit and helps me forget about things.",
    "I was sick at home.",
    "I needed a break.",
    "I needed to unwind.",
    "I had some downtime.",
    "I felt tired and wanted to relax.",
    "I was unhappy and wanted a distraction.",
    "I couldn't sleep so I scrolled for a bit."
]
},

```

```

"M7":{
"label": "Habitual Use",
"description": "Habitual social media use describes the development of an automatic, frequent behavior of engaging with social media platforms, driven by environmental cues and occurring with minimal awareness or intentionality. Therefore, the response states that social media was used automatically, out of habit, no clear motivation, unclear about the intention, or describing a repeated pattern of use should be labeled as habitual use. Note that the following cases should not be labeled as habit: ",
"examples": [
    "Just a habit.",
    "I have used it because it is an addictive force of habit and helps me forget about things.",
    "I used it because it is an addictive force of habit.",
    "I do not really know why i was on tiktok",
    "I open it without thinking much.",
    "I automatically opened it.",
    "No specific intention",
    "I did not realize I was scrolling",
    "I woke up and just started using it.",
    "I find IG being so funny so now I check it once or twice every day"
],
"exclusions": [
    " "I checked while I had some downtime." → Not Habitual Use; should be Passing Time (M5).",
    " "Just scrolling to pass some time." → Not Habitual Use; should be Passing Time (M5).",

```

" "I open TikTok because I got done doing what I was supposed to be doing early so I got bored and picked up my phone." → Not Habitual Use; should be Passing Time (M5)."

```
]
},
'M8': { "label": "Productivity / Learning",
  "description": "Using social media for instrumental, professional, or structured learning purposes. This includes (1) work, career, business, or productivity-related use (e.g., creating or sharing work content, networking, promoting business), and (2) educational or knowledge-building activities that involve learning about substantive domains (e.g., academic subjects, specific topics, or formal skills training). Unlike everyday information or checking casual updates, responses here explicitly reflect professional, academic, or knowledge-oriented goals.",
  "examples": [
    "I used it to create a WFH video.",
    "To network on LinkedIn.",
    "To follow a lecture on YouTube",
    "To post content for my business",
    "I wanted to research something about a vaccination.",
    "I was researching the subject of the Shah of Iran.",
    "To gain knowledge about a topic I'm studying."
  ]
},
```

```
'M9': { "label": "Trigger-Driven Use",
  "description": "A trigger that prompted social media use. This includes: 1) System-level triggers (e.g., push notifications, app updates, reminders), 2) Social triggers where another person's action prompted the response (e.g., a friend/coworker messaged, someone commented, being tagged).",
  "examples": [
    "got a notification",
    "friend sent me a message",
    "app asked me to update",
    "respond to a message",
    "read a message from my friend",
    'messaging family back',
    'my friend sent me a link'
  ]
},
```

```
"M10": {
  "label": "Activity/Context Only",
  "description": "This category is reserved for responses that only describe an action (watch video) or context, without giving any reason, purpose, or specific content focus. It applies only when no underlying motivation can be inferred. Do not use this category if the response identifies: (1) a type of content being sought (e.g., memes, billiard posts, local groups), (2) a
```

clear state or goal (e.g., relaxing, passing time, connecting, learning), or (3) a trigger such as responding to others.",

"examples": [

"I wanted to watch some videos.",

"To scroll TikTok.",

"To browse content.",

"Looking around.",

"Check timeline.",

"I went on the app.",

"Just looking at posts.",

"Watched videos.",

"Scrolling.",

"On Messenger."

],

"exclusions": [

"Mentions of a type of content being sought (e.g., 'memes,' 'billiard posts,' 'local groups') → should be Information Seeking.",

"Mentions of a clear state or goal (e.g., relaxing, passing time, connecting, learning) → should be Entertainment/Relaxation, Passing Time, or Coping, etc.",

"Mentions of responding to others (e.g., replying, messaging back, reacting to a friend's post) → should be Trigger-Driven Use."

]

}

}

```
def _get_prompt_classify_motivation(codebook_type_of_use, text):
```

```
    prompt = textwrap.dedent(f"""\
```

```
You are a research assistant helping with a thematic analysis of survey responses.
```

```
We want to understand people's motivations for using social media.
```

You are given:

1. A codebook of categories, with labels, descriptions, examples, and exclusions (for several categories).
2. A short survey response from a participant describing why they used social media in the past hour.

Your task:

- Read the participant response carefully.
- Assign one or more motivation categories from the codebook. A single response may belong to several categories.
- Only choose categories from the codebook. Do not invent new categories.
- If none of the categories apply, output exactly "Not-Applicable".

Output rules:

- If the response fits one category, output only that category label (e.g., "Information Seeking").
- If the response fits more than one category, output all applicable labels separated by a comma and a space (e.g., "Passing Time, Habitual Use").
- Do not output anything other than the label(s) or "Not-Applicable". Do not add explanations or extra text.

Codebook:

{codebook_type_of_use}

Survey response:

"{text}"

Now please output the assigned label(s):

""")

return prompt

Chapter 4 – Digital well-being features on social media platforms: an LLM-assisted analysis of social media company announcement

Authors

Yuning Liu¹, Stanley Huang², Emily Xing³, Ravi Iyer⁴, Kasisomayajula (Vish) Viswanath¹

Affiliations

¹ Harvard T.H. Chan School of Public Health & Dana-Farber Cancer Institute

² Boston University

³ Harvard University

⁴ Neely Center for Ethical Leadership and Decision Making, Marshall School of Business, University of South California, Psychology of Technology Institute

Abstract

Digital media platforms increasingly face pressure from the publics and legislators worldwide to mitigate harm to user well-being and promote healthier digital environments. Platforms' response has been to proclaim actions in response to public concerns. This study examines platforms' claims to promote digital well-being practices through their public company announcements and analyzes these practices through the lens of a harm-responsive design process. This study provides arguably the first large-scale analysis of digital well-being related practices claimed by platforms using a large language model assisted content analysis of 13,497 company announcements published by seven major platforms (Facebook, Instagram, WhatsApp, X/Twitter, TikTok, YouTube, and Snapchat) from each platform's first post through May 15, 2025. We identified 5,685 distinct claims about digital well-being practices claimed by the platforms and characterized their distribution across platforms, target user groups, harm categories, intervention mechanisms, and whether the impact of these practices are being assessed. Our findings show

that platforms have reported the adoption of a growing number of digital well-being practices since 2019, with a surge following the introduction of the EU General Data Protection Regulation. Approximately one third of all practices focused on harmful content, followed by harmful contact and user privacy, while relatively few addressed harmful user behaviors such as excessive scrolling. Across the four identified mechanisms of digital well-being practices: content-based mechanisms, user-initiated experience control, default access and account protection, and personalized algorithm curation, more than 80% fall within content-based governance and user-initiated controls, with far fewer addressing default safety features or personalized algorithmic curation. Only a small share of announcements includes user-level impact evaluation, such as how many users were affected or whether the intended outcomes were achieved, leaving the implementation status and actual effects of these practices largely unknown. Platforms appear responsive to public and legislative pressures, but the actions they highlight tend to emphasize visible, content-focused measures rather than deeper changes to underlying algorithmic operations. These findings reveal substantial transparency gaps in platform-reported digital well-being practices, including limited information about what platforms actually implement and a lack of evidence regarding the user-level impact of these interventions. We call for harm-responsive, evidence-based, and globally coordinated systems for documenting platform design choices and their impacts.

4.1 Introduction

The simultaneous rise in social media use, negative online experiences, and mental health crises, particularly among youth and adolescents, has become a major global concern. Digital well-being, defined as the impact of technologies and digital services on mental, physical, social, and emotional health [216], represents one of the most pressing challenges of our time.

In response to these concerns, a range of digital self-control tools has been developed [199]. These tools are delivered as web extensions or mobile applications [217] and grounded in

theories of user agency [218], self-regulation [219], and behavioral design such as the HOOK model [220]. These tools implement interventions such as blocking or content removal, self-tracking, goal advancement, and reward or punishment mechanisms [199,217]. However, external digital well-being tools require users, and often parents, to be aware of them, able to afford them, and possess the skills and motivations to use them consistently, creating major barriers to their widespread adoption and sustainable long-term use, and may exacerbate inequality. In other words, most digital self-control tools are “self-programmed” by the user community and heavily rely on “self-regulation” with continuous motivation of individual users [217].

Recognizing that digital well-being cannot rely solely on individuals, governments and legislators have begun introducing laws that hold social media platforms responsible for creating safer and healthier online spaces. As shown in Table 3.1, regulatory efforts addressing digital well-being are emerging globally, with early frameworks in the EU and UK, followed by expanding initiatives in countries such as Australia, Brazil, India, Rwanda, China, and the United States. With the growing global momentum around digital well-being legislation, it is worth examining the response of the platforms with avowed claims about mitigating any harmful impacts. By analyzing platform-level actions, researchers and policymakers can assess regulatory gaps, evaluate the effectiveness of digital well-being initiatives, and guide future governance and design strategies.

This study examines how social media platforms responded to these external pressures by claiming to alter their platforms’ features to promote digital well-being practices. In this study, we use the term digital well-being practices to refer to the design features, policies, and initiatives introduced by social media platforms to reduce user harm, address mental health concerns, enhance positive experiences, and promote healthier technology use.

Table 3.1. Selected examples of policies addressing digital well-being and online safety

Policy Name	Jurisdiction	Year Adopted	Effective Date	Target Population	Regulatory Focus
General Data Protection Regulation (GDPR)	EU, UK	2016	May 25, 2018	All users	Strengthening user control over personal data through limits on data processing and consent requirements
Digital Services Act (DSA)	EU	2020	Feb 17, 2024	All users	Imposing risk-based transparency and accountability requirements on online platforms
Online Safety Act	UK	2023	July 25, 2025	All users	Mandating platform-level child safety, content moderation, and reporting obligations
Social Media Minimum Age Law	Australia	2025	Dec 10, 2025	Children under 16	Restricting social media access for minors to protect child mental health and well-being
Digital Personal Data Protection Act	India	2023	2023	All users	Regulating personal data collection and use through consent-based data protection rules
Online Child Protection Policy & Ministerial Instruction	Rwanda	2019 / 2024	2019 / 2024	Children and adolescents	Mandating content filtering, age verification, and, content moderation for providers
Resolution No. 245	Brazil	2024	2024	Children and adolescents	Defining children's rights in digital environments, including privacy and access to information
Regulation on the Protection of Minors Online	China	2023	Jan 1, 2024	Minors	Imposing platform responsibilities through social co-governance to protect minors' well-being and digital rights
Age-Appropriate Design Code Act (AADC)	United States (California)	2022	—	Children and adolescents	Requiring age-appropriate design and default settings to reduce risks to minors
Social Media Regulation Act	United States (Utah)	2023	—	Minors	Introducing age verification, parental consent, and limits on addictive platform features

Study platforms' digital well-being practices through company announcements

To examine how social media platforms have implemented internal practices for digital well-being, we analyzed company announcements published through official newsrooms or websites. Two recent reports have reviewed digital platforms' company announcements related to child safety [221,222]. One report identified 91 platform changes influenced by the Age-Appropriate Design Code Act (AADC) between 2018 and 2022, covering four categories: age-appropriate design, time management, privacy and data management, and youth safety and well-being [221]. Another study identified 128 platform changes related to child privacy and safety from 2017 to 2024, grouped into four categories: defaults, tools, information, and support for adolescent users [222]. However, no study has yet systematically examined digital well-being design features across different user populations, including how platforms describe, justify, and promote these features. Such evidence is essential for answering the core question of what social media platforms have actually done to promote digital well-being.

We focus on company announcements because our central question is what platforms publicly claim to have done in response to growing external pressure from users, advocates, and regulators. Company announcements are the primary vehicle through which platforms communicate these claimed responses, making them directly available evidence of how platforms frame and justify their actions on digital well-being. By analyzing the announcements, we take platforms at their own words, examining what they say they have changed and how they characterize the problem they are addressing, which is itself informative for questions of regulatory oversight and public accountability. Moreover, other potential data sources are also limited. External researchers lack access to internal company documents. Existing resources such as the Facebook Archive cover only a single platform and provide partial information within a time period. While corporate transparency reports offer some quantitative metrics of platform governance activities, they lack the rich contextual meaning and justification of platform actions that company announcements provide.

Examining digital well-being practices through a harm-responsive design process

Beyond the need for a comprehensive overview of design changes, platform digital well-being practices require a structured approach to assess their impact on user well-being. According to the taxonomy of social media harms [223], key issues of social media on user well-being fall into four categories:

- Harmful or unwanted content: exposure to potentially damaging content such as misinformation, violence, nudity, and extreme material;
- Harmful or unwanted user behaviors: users developing negative or harmful psychological reactions to social media use, such as negative social comparison and fear of missing out; and users develop in response to social media use, such as loss of self-control and problematic use;
- Harmful or unwanted contacts: harmful interactions from other users or bots, such as bullying or unwanted sexual advances;
- Privacy violations: threats to users' data and accounts on social media platforms, including privacy invasion, impersonation, and exposure to malware or adware.

More importantly, specific design elements contribute to these harms. Features like infinite scroll, ephemeral content, and push notifications may encourage excessive use, while discriminatory algorithms, attention optimization, and the lack of negative feedback channels could amplify harmful content [224].

Given the link between platforms' structural designs and harms, we evaluated whether platforms follow a systematic harm-responsive approach in developing and assessing digital well-being interventions (Figure A3.1). The harm-responsive approach consists of three key steps: harm identification, where platforms assess harms and the design choices motivated by them; design changes implemented to mitigate these harms; and impact evaluation, where the

outcomes and effectiveness of these interventions are systematically assessed. In summary, this study focused on two research questions:

- R1: Evaluating the coverage, content, and design of digital well-being initiatives implemented by social media platforms through company announcements.
- R2: Determine whether platforms explicitly claim that their digital well-being initiatives address identifiable well-being concerns and whether they report evaluations of their effectiveness.

4.2 Method

The data for this study come from an analysis of 9,542 company announcements from the websites of seven social media platforms. We examined the content of these announcements to address the research questions with the assistance of Large Language Models (LLMs) [225,226].

4.2.1 Data Collection and Preprocessing

We collected all company announcements and news articles (N=13,497) published on the official newsrooms of seven social media platforms, including TikTok, Meta (Instagram, Facebook, and WhatsApp), YouTube, Snapchat, and Twitter/X, from the earliest available date up to May 15, 2025. These platforms were selected because they are among the largest social media platforms by daily active users and are primarily smartphone-based [227]. The source websites for the seven platforms are listed in Table A3.1. We then applied a keyword dictionary–based search to identify articles relevant to digital well-being. To be included, a company announcement had to contain at least one keyword related to well-being, user experience, youth, or other related themes, as detailed in Table A3.2. After keyword filtering, 9,542 company announcements remained for analysis.

We also removed duplicate company announcements that appeared in different languages or were reposted across regional newsroom sites. For example, TikTok often republished identical announcements through multiple country offices or language versions. To identify duplicates, we employed a two-step process: (1) we compared URLs and titles, as many duplicates shared nearly identical titles (e.g., `teens_update_2020q1`), and (2) we embedded text with Sentence-BERT, a sentence-level transformer model that encodes text into semantic vector representations, and calculated cosine similarity scores between pairs of TikTok announcements published within a one-month window [228]. For non-English items, we used GPT-4o to translate all texts into English before analyses. Announcements with a cosine similarity above 0.7 were treated as duplicates and merged into a single entry. After removing duplicates, the final dataset comprised 7,318 unique company announcements.

4.2.2 Codebook Development

An initial codebook was developed by the investigators through the review of 100 randomly selected articles and informed by a literature review of relevant concepts and categories to be extracted. The final codebook, detailed in the supplementary materials, includes both deductive and inductive dimensions. The deductive codes were derived from theory and prior literature to capture key aspects of digital well-being practices, including demographics (target population or user group), geography (scope or implementation region), harms (type of user harm addressed, following the Social Media Harm Taxonomy), design feature (platform or design feature involved, following the Platform Design Taxonomy), motivation of platforms (stated reason or driver for the practice), automaticity (whether the feature is default or user-initiated), instruction (information or guidance provided to users), evaluation plan (approach to assessing effectiveness), impact (reported or intended outcome), and partnership (external collaborators or affiliated entities). The inductive codes, derived from iterative reading and thematic analysis,

include design features (specific platform mechanisms identified in text) and specific harms (concrete harms described beyond predefined categories).

4.2.3 Human coder labeling

With the codebook developed in 4.2.2, two coders underwent multiple rounds of training to familiarize themselves with the coding framework and ensure consistent application. Content coding was conducted in Qualtrics following the established codebook. To assess initial consistency, the coders independently annotated 10 randomly selected news articles, sampled to include at least one from each social media platform. Inter-coder reliability was then assessed using ICC. For items that demonstrated low agreement, the research team met to discuss discrepancies, refine code definitions, and clarify coding instructions. The coders subsequently coded an additional 40 randomly selected articles, and reliability was reassessed. After two iterations, satisfactory ICC levels were achieved (see Table A3.3). Each coder then independently coded 75 additional articles, resulting in a gold-standard labeled dataset of 200 articles, which served as the reference for evaluating LLMs performance against human annotation.

4.2.4 LLM-Assisted Pipeline for Deductive Content Classification

Given the large volume of platform company announcements with design changes, we applied LLMs to assist in content analysis. LLMs have been widely used in text classification, named entity recognition, summarization, and both inductive and deductive coding. They have proven effective in psychological text analysis [229], deductive coding using predefined codebooks [225], and inductive thematic analysis [226,230]. In this study, we extract key information from platform-disclosed digital well-being practices and apply LLMs to support qualitative content analyses of company announcements across multiple dimensions. One core challenge in this content analysis lies in the mismatch between the sampling unit (company announcements) and the analytical unit (individual digital well-being practices). A single

announcement may describe multiple design practices, each of which constitutes a distinct analytical unit. Consequently, extracting precise information for each analytical unit requires addressing two major linguistic challenges: long-range dependencies, where relevant concepts persist across multiple sections of an article, and co-reference complexity, where a single sentence may refer to multiple concepts. To address these challenges, we developed an LLM-assisted content analysis pipeline that systematically captures and organizes information across such dependencies. The overall workflow is illustrated in Figure A3.2, with detailed prompt provided in the supplementary materials and prompt structures described in Figure A3.3.

Our pipeline employs a multi-stage, LLM-assisted approach. First, each article is unitized or decomposed into discrete, self-contained references to the analytical unit, the digital well-being design practices, based on predefined inclusion criteria (see Figure A3.4). To qualify as a digital well-being practice, (1) the practice must be explicitly associated with reducing digital harms or promoting digital well-being, and (2) it must involve a clearly defined and implemented product or feature integrated into the platform’s design (see supplementary materials, Prompts, Step 1). Next, after pooling all analytic units, the model references the original article to annotate each unit according to an annotation schema (see Figure A3.5). Finally, classification is performed directly on these annotated analytic units.

Following best practices in LLM-assisted content analysis [225,231–234], we designed a structured prompting framework that employed several complementary techniques: a few-shot prompting approach leveraging in-context learning to minimize task-specific fine-tuning; persona descriptions to align the model with domain-specific expertise; chain-of-thought reasoning for step-by-step analysis and error inspection; and structured output specifications in JSON format to standardize results. All classification tasks were reformulated as binary classification problems. We used GPT-4o (temperature = 0, n = 5 repetitions) to simulate independent coders, following prior work on self-consistency [235,236]. We combined a majority-vote procedure with perplexity-based selection to determine the most representative response among the five repetitions for

downstream analysis. The model was also instructed to estimate its uncertainty level; cases with high uncertainty were flagged for human review. To assess robustness, we manually inspected 50 company announcements to verify that the model did not hallucinate or misattribute information when generating abstractive summaries and extracting textual references for each platform digital well-being practice. We compared the machine-coded responses with the ground truth generated by human research assistants (Table A3.3).

4.2.5 LLM-Assisted Thematic Content Analysis

We also applied LLM to assist the thematic analysis on platform design features and targeted harms described in company announcements. The process followed a three-step, human-in-the-loop framework integrating qualitative reasoning from researchers with model-based scalability. First, we prompted the LLM to extract a concise action code from each digital well-being practice. Each code represented a concrete, platform-level design intervention or initiative, expressed as a short, neutral phrase describing what the social media platform was doing or implementing. Then, we instructed the LLM model to synthesize and merge the initial action codes into broader, non-overlapping design themes through iterations over all company announcements. The model was instructed to generate between 15 and 30 well-defined themes, ensuring interpretability and parsimony, while placing ambiguous items in an “Uncategorized” group. Human researchers then reviewed and refined the merged codebook, resolving redundancy and ensuring conceptual coherence. Finally, the LLM applied the revised theme codebook to classify the full corpus, assigning each digital well-being practice to one or more design categories. Details of the prompt are provided in the Prompts section in the supplementary materials, Step 4. For the text classification step, we evaluated the model’s performance by comparing its coding results with 100 human-coded practices based on the revised codebook.

4.3 Results

From the 13,497 company announcements collected across Meta (Facebook, Instagram, WhatsApp), TikTok, YouTube, Snapchat, and Twitter/X, a total of 7,318 unique announcements remained after preprocessing and duplicate removal. Applying our coding pipeline to this dataset, we identified 3,175 announcements related to digital well-being. From these, we extracted 5,685 distinct digital well-being practices introduced from as early as 2005 to the end of our data collection period on May 15, 2025. Among the identified digital-wellbeing practices, TikTok accounted for the largest share ($n = 3,433$; 60.39%), followed by Facebook ($n = 705$; 12.40%), Instagram ($n = 488$; 8.58%), X/Twitter ($n = 323$; 5.68%), YouTube ($n = 316$; 5.56%), Snapchat ($n = 290$; 5.10%), and WhatsApp ($n = 130$; 2.29%).

4.3.1 *Digital well-being practices by platforms*

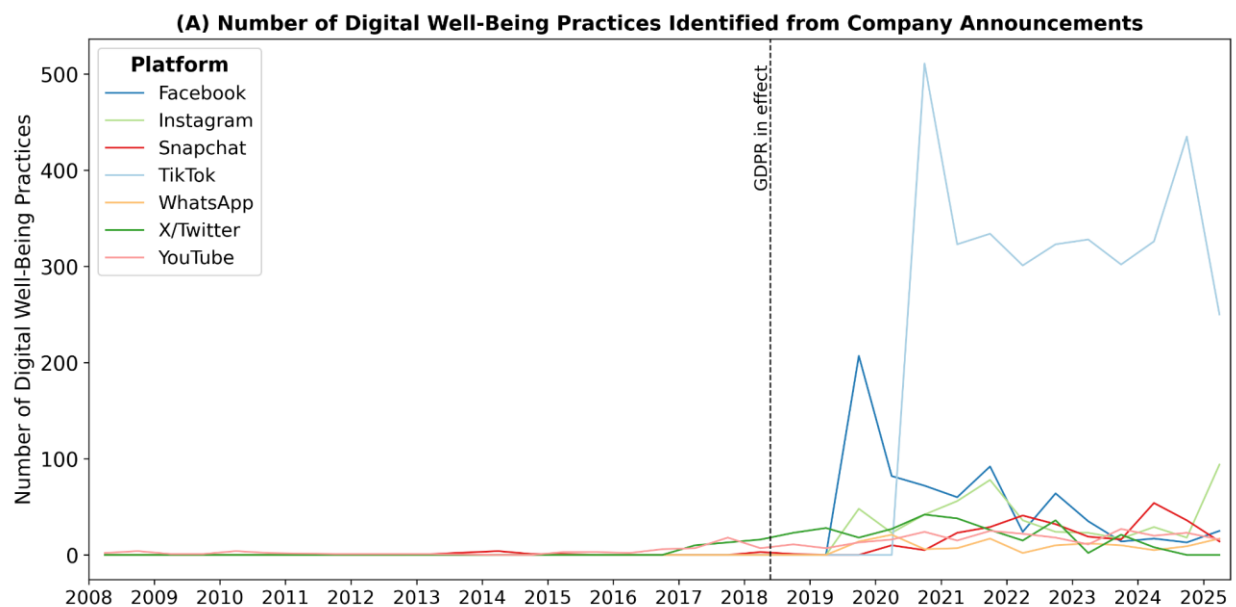
All platform-initiated digital well-being designs began to increase noticeably after 2019. As shown in Figure 3.1 panel A, Facebook and Instagram exhibited the first major waves of reactions starting in 2019, followed by a sharp rise in TikTok's digital well-being practices in 2020. Among the 511 digital well-being practices on TikTok in July-December 2020, 62% were related to the dissemination of health-related information or content moderation, as will be detailed below.

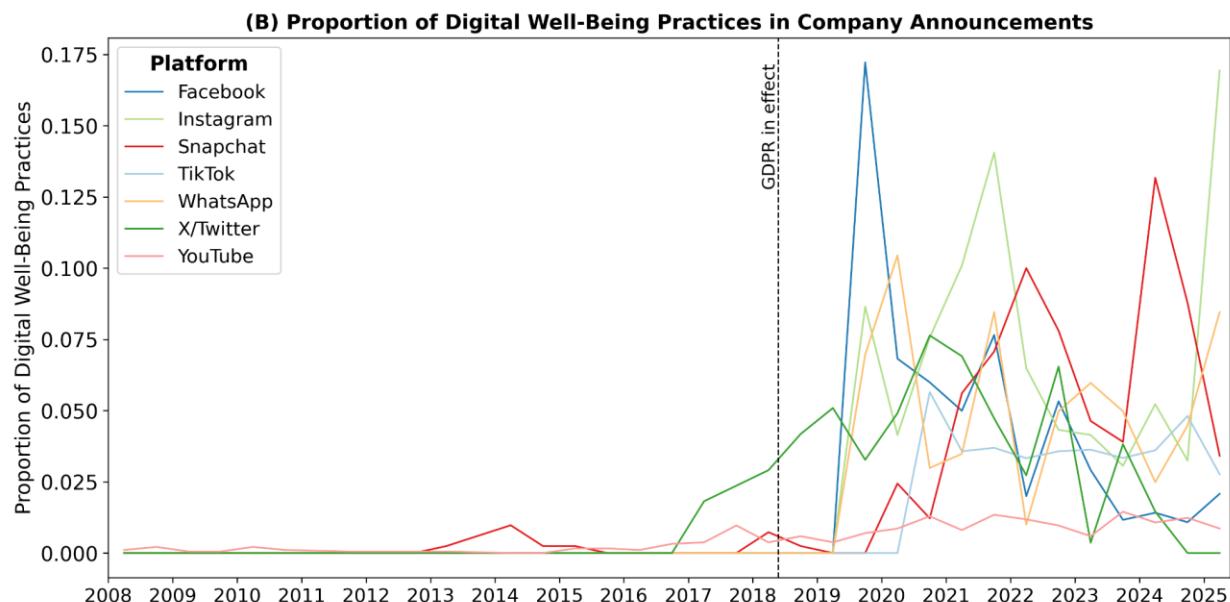
Notably, multiple platforms exhibit a co-increasing trend in digital well-being announcements within a similar time window beginning in 2019. This coordinated shift temporally aligns with the implementation of the European Union's General Data Protection Regulation (GDPR) [23] in 2018 (the dotted line in Figure 3.1), suggesting that rising regulatory pressure may have catalyzed broader signaling efforts across platforms.

When comparing the proportion of digital well-being announcements relative to all company announcements, as shown in Figure 3.1 panel B, Facebook shows the first and the highest increase, rising from nearly zero to a peak of 18% in July-December 2019. Instagram and WhatsApp followed similar trends, increasing from nearly zero to 10% and 8%, respectively,

during 2019. A second surge in the proportion of company announcements with digital well-being practices occurred in 2021–2022, during which Instagram’s proportion rose to 14% and Snapchat’s to 10%. Twitter/X displayed a steady upward trend in digital well-being practices from 2017 to 2022 (ranging between 5–8%) but experienced a decline after 2023 after the reorganization. Although TikTok reported a large absolute number of digital well-being initiatives, its overall proportion remained below 7% when adjusted for its high total volume of company announcements.

Figure 3.1. Introductions of digital wellbeing practices for platforms from 2008-2025



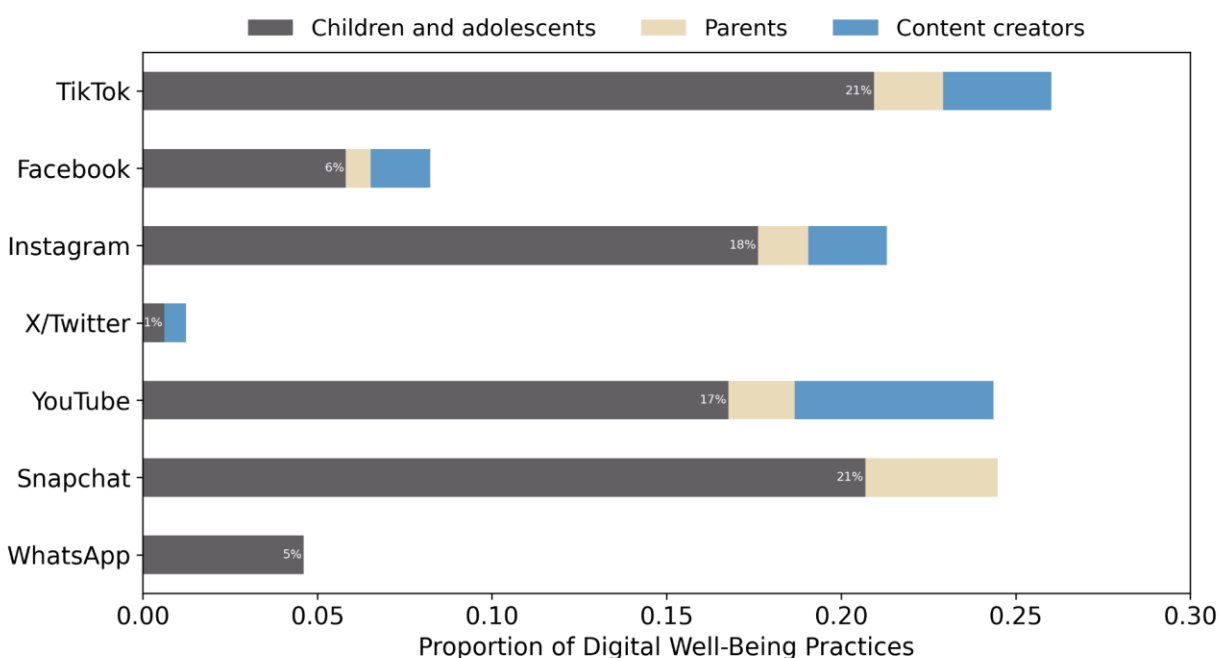


Note: The plot presents, for each platform and half-year period, (a) the raw number of platform digital well-being designs, and (b) the proportion of digital well-being practices relative to the total number of company announcements extracted from official websites. Data were collected from official news channels where social media platforms publish company announcements and updates.

4.3.2 Target users of digital well-being practices

We observed that TikTok and Snapchat have the highest proportions of digital well-being designs aimed at addressing harmful effects on children and adolescents (both at 21%), likely reflecting the larger share of young users on these platforms, shown in Figure A3.2. Across all platforms, less than 5% of practices are directed toward parents, with Snapchat showing the highest proportion within this group. Meanwhile, TikTok and YouTube report 3% and 5% of practices, respectively, targeting content creators, partly reflecting their creator-centric cultures and platform structures.

Figure 3.2. Target user groups of digital well-being practices by platforms



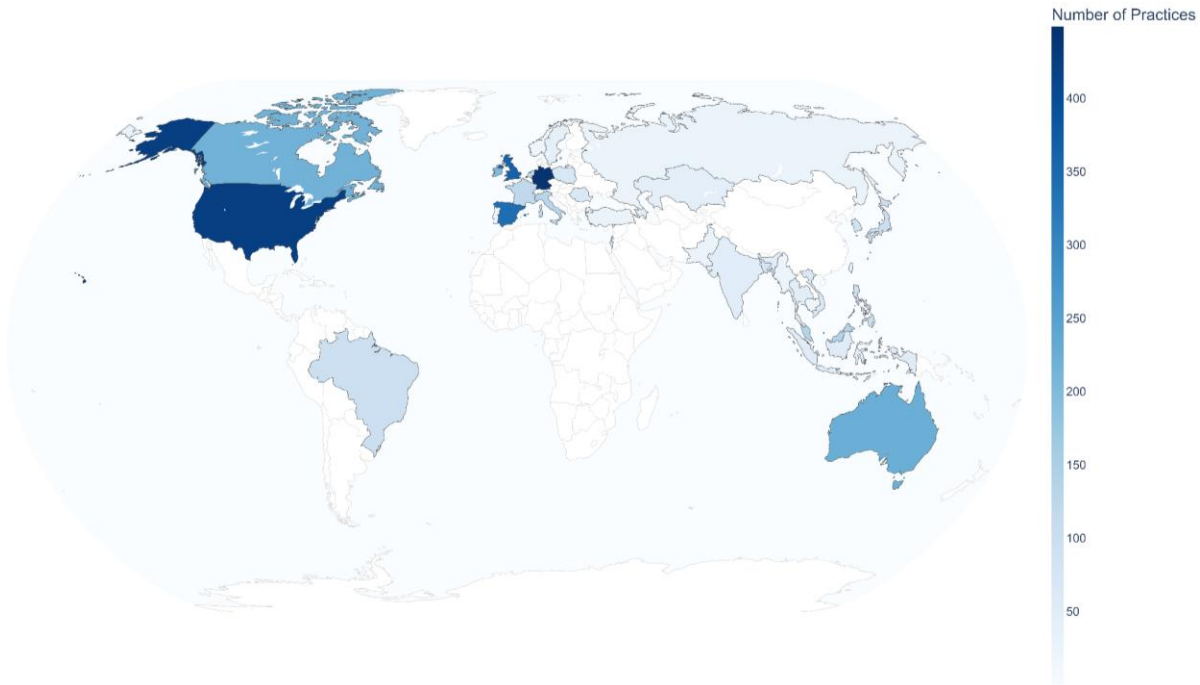
Note: The chart illustrates the primary target users of each platform's digital well-being practices, that is, the intended audience specified in the company announcements. When the target audience was not explicitly stated, we assumed the practice was directed toward general users (not shown here). Practices explicitly target children and adolescents, parents, or content creators are displayed. Figure A3.6 presents the raw number of digital well-being practices categorized by platform and target user.

4.3.3 Geographic distribution of digital well-being practices

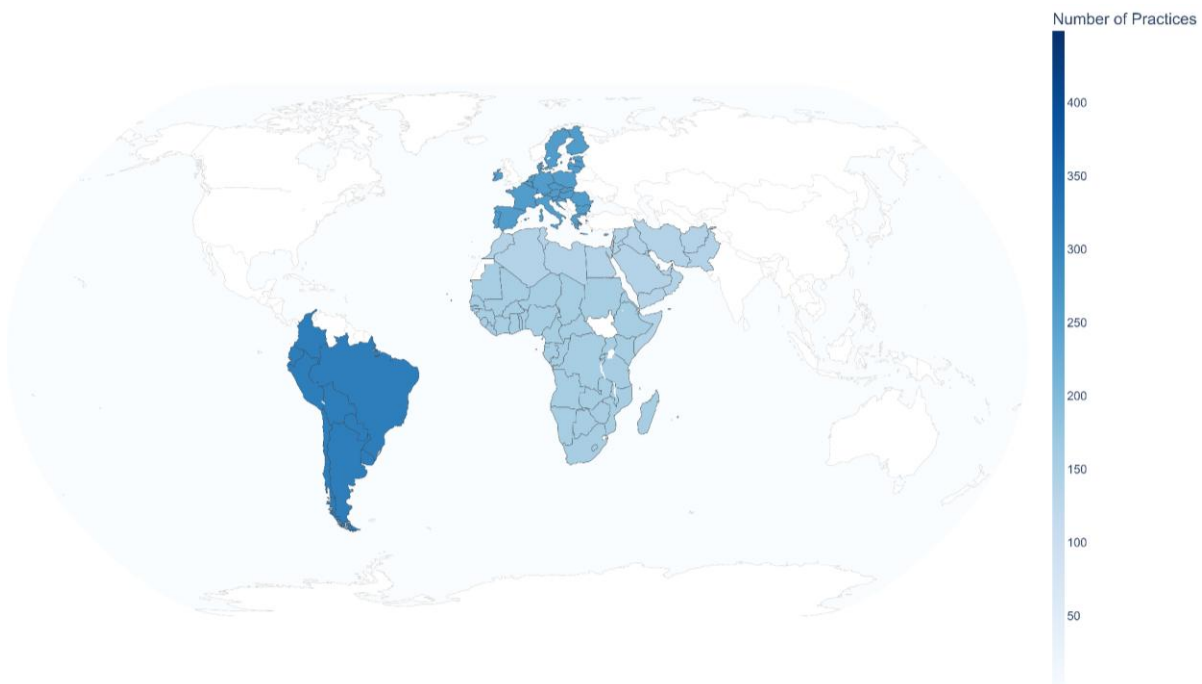
Germany, the United States, Spain, the United Kingdom, and Australia are the leading country offices posting digital well-being practices on TikTok, followed by other European offices such as France and Italy, and Asian offices including Japan, Malaysia, Brazil, and India (Figure 3.3). At the regional level, the Latin America office published the largest number of digital well-being posts, followed by the European, African, and Middle East & North Africa offices. Notably, the color scales in panels A and B are comparable, indicating that the volume of digital well-being posts from Germany and the United States exceeds the total number of posts from any regional office. This pattern may suggest a systematic bias in platform responsiveness: TikTok appears more active in communicating digital well-being commitments in Global North markets than in Global South, where regulatory pressure is stronger and enforcement mechanisms more developed.

Figure 3.3 Distribution of digital well-being practices in TikTok

(A) By TikTok country office



(B) By TikTok regional office



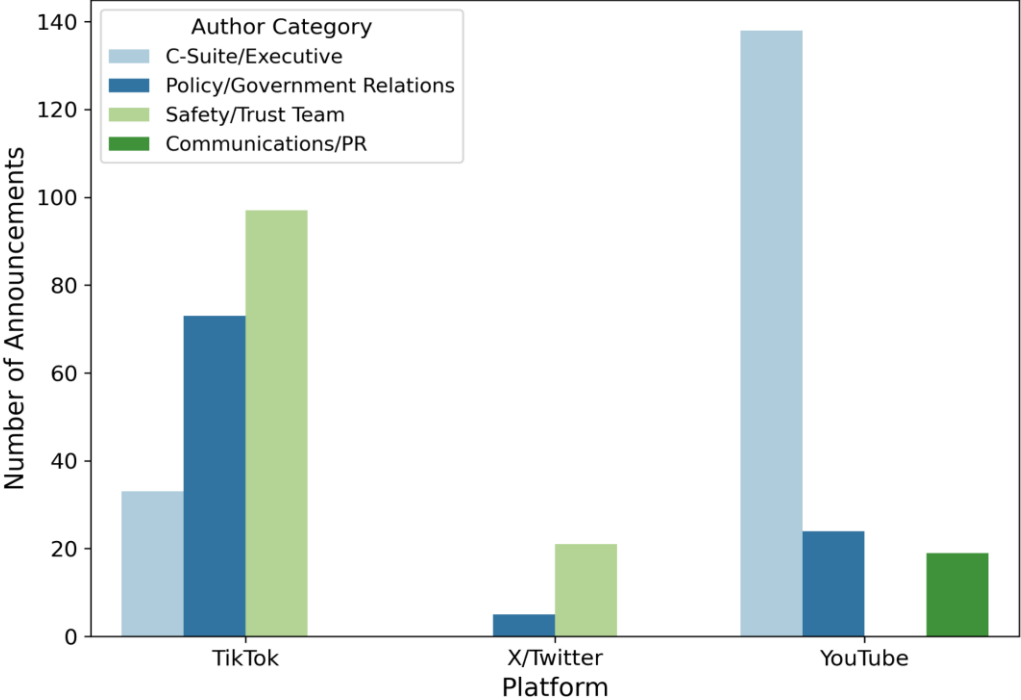
Note: Among the platforms included in this study, only TikTok released company news announcements through country-specific channels, allowing us to accurately determine the geographic distribution of its digital well-being practices. Other platforms, such as Meta, occasionally referenced regional or country-level activities in their announcements, but these mentions were inconsistent and incomplete. Therefore, we report detailed geographic results only for TikTok. TikTok operates both country-level and regional offices. Accordingly, panel (A) presents the number of practices reported by each country office, while panel

(B) presents the number of practices released by three regional offices, including the European Union, Middle East & North Africa, Africa, and Latin America. Because TikTok’s announcements did not explicitly define the countries covered by each regional office, we inferred regional groupings based on the World Bank.

4.3.4 Authors of the digital well-being practices company posts

Most digital well-being practices were posted by one of the following groups: C-suite executives, policy and government relations teams, trust and safety teams, or communications teams. YouTube had a higher proportion of posts from C-suite executives, whereas TikTok and X/Twitter had more posts authored by their trust and safety teams.

Figure 3.4. Author positions in the company announcements with digital well-being practices



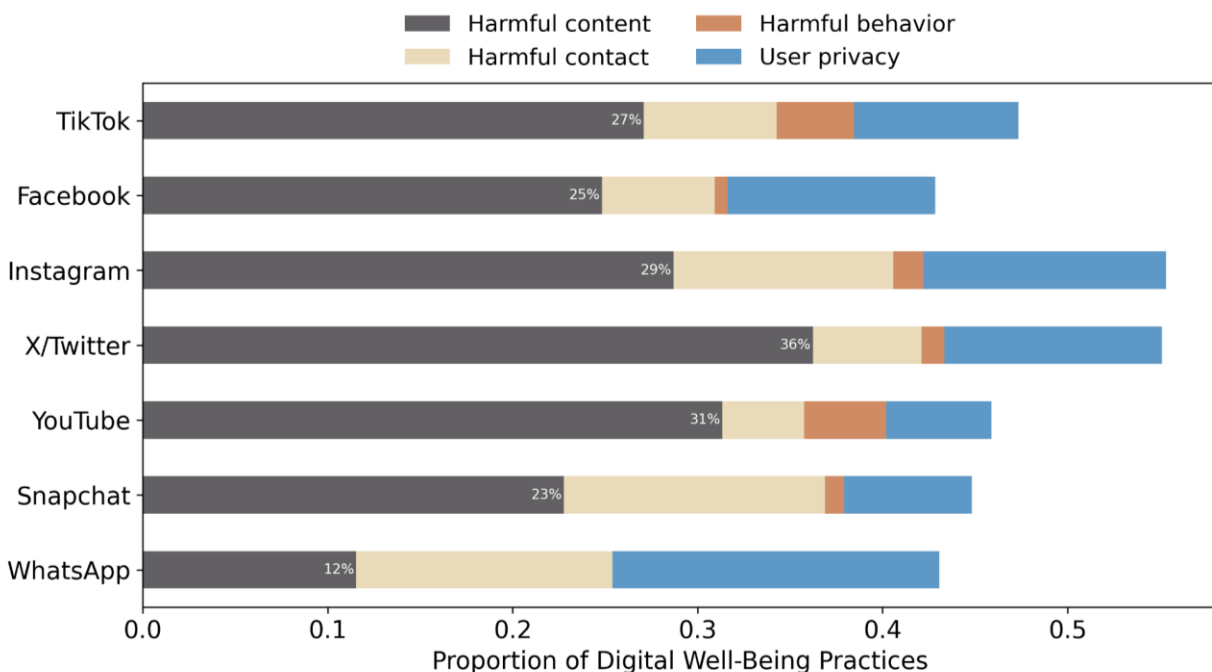
Note: Data were collected from official news channels where social media platforms publish company announcements. Among all platforms, only TikTok, X/Twitter, and YouTube specify author names or teams in their company announcements. Accordingly, we report data for these three platforms.

4.3.5 User harms targeted by digital well-being practices

The largest proportion of digital well-being practices across platforms, except for WhatsApp, focused on reducing harmful content through measures such as content moderation

and labeling, ranging from 23% on Snapchat to 36% on X/Twitter, shown in Figure 3.5. Following the social media harm taxonomy, harmful content refers to users' exposure to potentially damaging materials such as misinformation, violence, nudity, and other extreme content. User privacy harm, such as threats to users' data and accounts including privacy violations, impersonation, and exposure to malware or adware, constitutes another major category, accounting for 6% (YouTube) to 18% (WhatsApp) of practices across platforms. Harmful contact, defined as harmful interactions from other users or bots such as bullying or unwanted sexual advances, is most frequently addressed by Snapchat (14%) and WhatsApp (14%), consistent with the social and interactional nature of these platforms. By contrast, harmful usage, users' negative psychological or behavioral reactions to social media, such as fear of missing out, social comparison, loss of self-control, or problematic use, was the least frequently targeted category across all platforms.

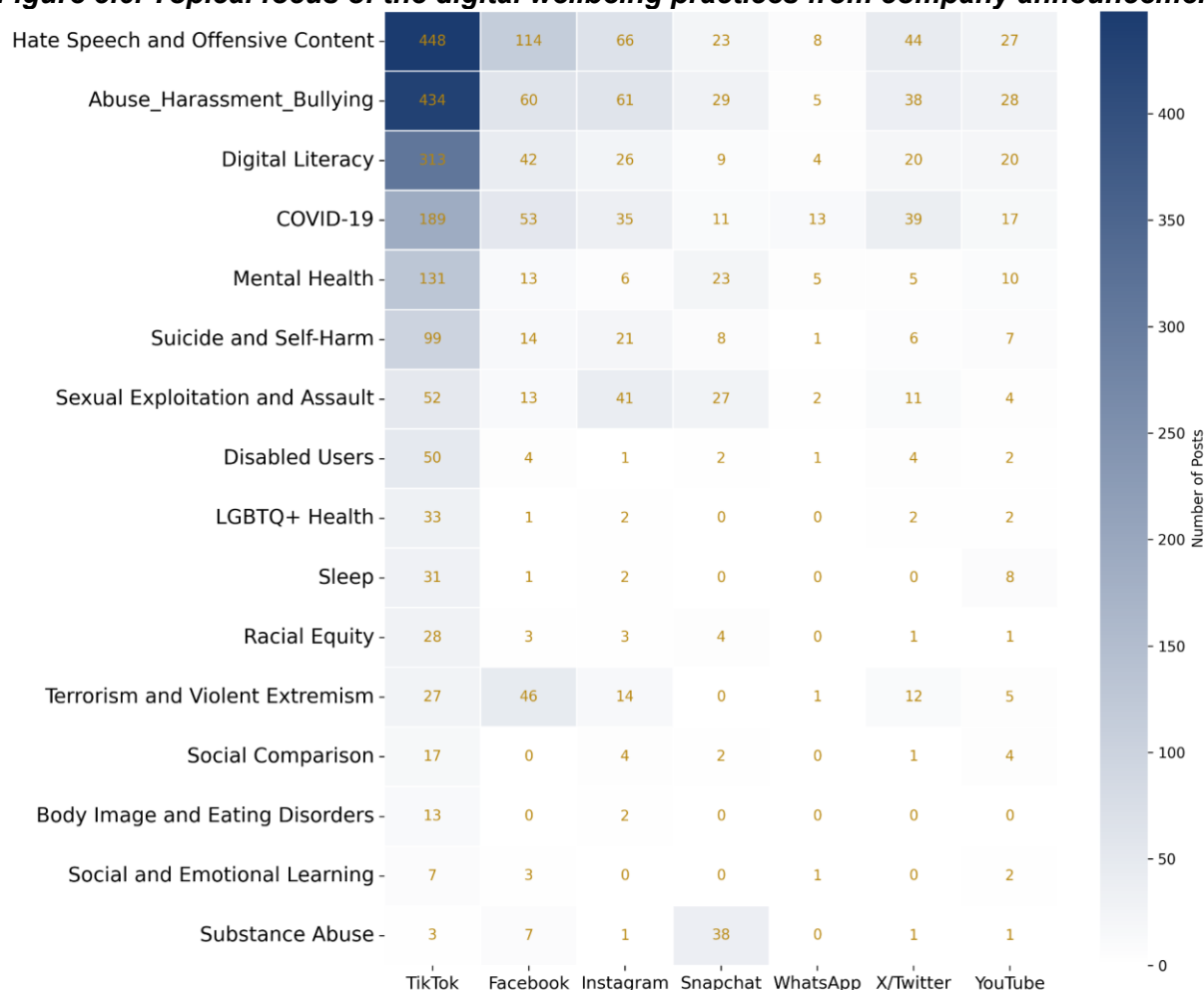
Figure 3.5. Targeted harms of the digital well-being practices in company announcement



Note: The plot illustrates the primary target harms addressed by each platform's digital well-being practices, categorized into four types based on the social media harm taxonomy: harmful content, referring to exposure to damaging materials such as misinformation, violence, nudity, or extreme content; harmful behavior, denoting negative psychological or behavioral reactions such as fear of missing out, social

comparison, or loss of self-control; harmful contact, involving harmful interactions from other users or bots, including bullying or unwanted sexual advances; and harmful use of information (user privacy), concerning risks to user data and accounts, such as privacy breaches, impersonation, or malware exposure. Some digital well-being practices do not fall into any of these categories. For instance, those focused on providing health or positive information which are not the scope of this paper. Figure A3.7 presents the raw number of digital well-being practices categorized by target harms.

Figure 3.6. Topical focus of the digital wellbeing practices from company announcements



Note: the topics were derived from the thematic content analysis of social media company announcements. Data were collected from official news channels where social media platforms publish company announcements.

Shown in Figure 3.6, across TikTok, Facebook, Instagram, and YouTube, hate speech and offensive content are among the most frequently mentioned topics in company announcements related to digital well-being practices. For instance, TikTok revised and expanded its definition of hate speech within its community guidelines [237]. In another post, Meta reported

substantial progress in detecting and removing such content in its 2020 Community Standards Enforcement Report [238], with the proactive detection rate for hate speech increased from 89% to 95% on Facebook, and rose from 45% to 84% on Instagram.

These efforts are followed by practices targeting abuse, harassment, and bullying. For example, in its post “Demonstrating Respect and Empathy on World Kindness Day,” Snapchat described measures to combat abuse and harassment by enforcing strict bans on non-consensual intimate imagery and sexual harassment [239]. The platform also offers in-app tools for users to block and report offenders, supported by its 24/7 Trust and Safety teams that review reports and take actions such as issuing warnings, suspending, or terminating accounts. Similarly, Meta introduced updates to its safety standards in Facebook community chats, giving group administrators and moderators enhanced control tools to maintain a safe environment [240].

Another recurring theme is digital literacy. TikTok, for example, provides users with contextual tools to assess content credibility, including labels for unsubstantiated videos and verification check marks to confirm account authenticity. It also applies state-affiliated media labels to identify government-influenced content, expanding this policy globally to promote transparency about information sources [241]. Moreover, topics related to mental health awareness and suicide prevention also appear in digital well-being practices. For instance, TikTok offers search interventions that direct users to local helplines such as Crisis Text Line for immediate support [242], alongside well-being guides and awareness campaigns like Mental Health Awareness Month [243].

4.3.6 Mechanisms of digital well-being practices

We also analyzed the mechanisms of digital well-being practices, defined as the specific actions, design features, and operational levers through which platforms implement them, encompassing both interface-level interventions and informational or governance measures. We identified four broad mechanism categories: content-based mechanisms, user-initiated

experience controls, default data safety and account protection, and personalized algorithmic content curation.

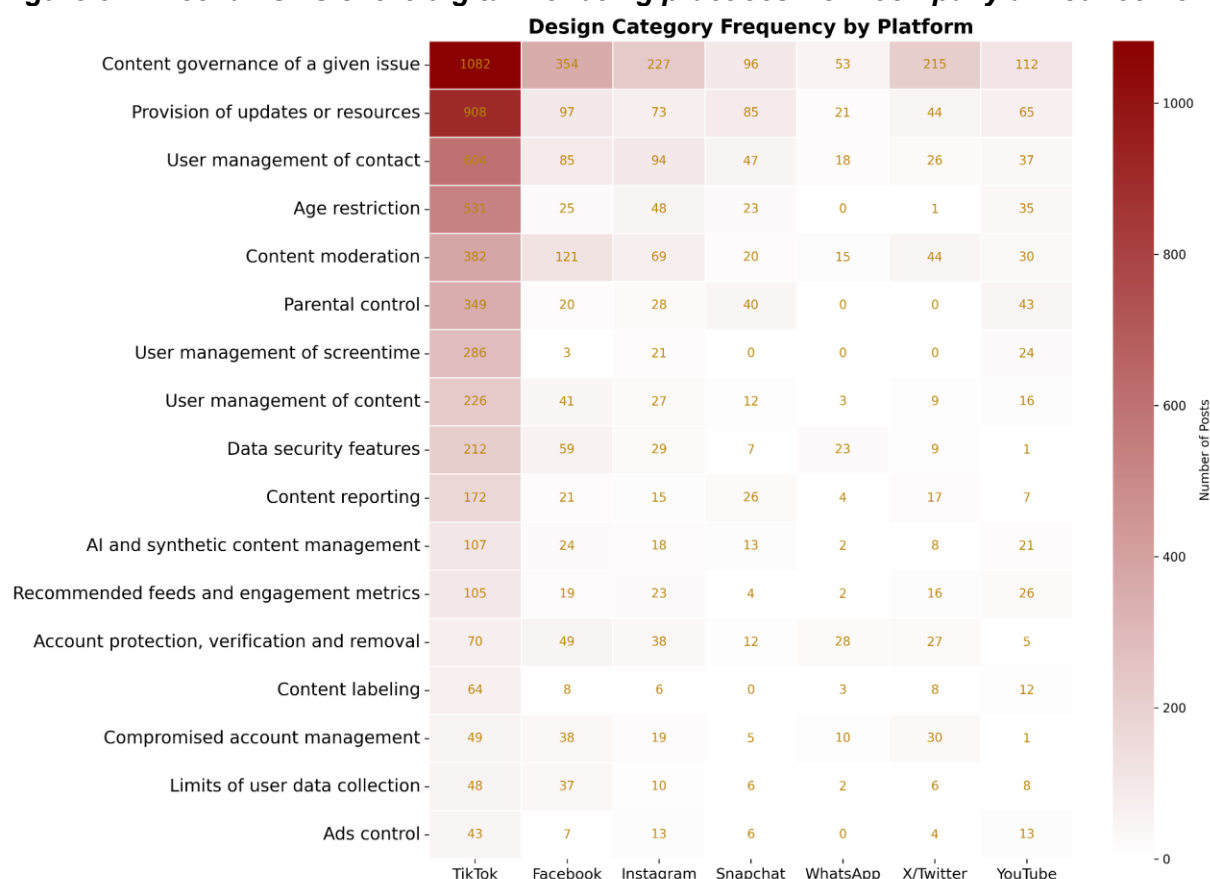
Content-based mechanisms

The analysis of platform-initiated digital well-being mechanisms reveals a clear concentration of practices at the content level. As shown in Figure 3.7, the most prevalent mechanisms include the provision of updates or resources, content governance on specific issues, and content moderation. Regarding the provision of updates or resources, platforms often deliver educational, civic, and health-related information. For instance, in April 2021, TikTok [244] partnered with experts and organizations to provide educational resources on sexual assault awareness, featuring learning content on its Discover page and permanent public service announcements linking to help resources in its Safety Center. Similarly, Instagram [245] began surfacing localized eating disorder hotlines and expert advice when users searched for, shared, or encountered disordered eating content in February 2021, and added options for users to message friends directly from these in-app resources for timely support.

Another content-focused digital well-being mechanism highlighted in company announcements is issue-based content governance. While it sometimes overlaps with information and resource provision, this category centers on platform-led initiatives aimed at addressing specific societal events or crises, such as public health emergencies, elections, or misinformation surges, to maintain the integrity and reliability of online information. For example, during the COVID-19 pandemic, TikTok [246] provided verified public health information by hosting an in-app WHO information hub, broadcasting live sessions with WHO experts, and labeling millions of COVID-related videos with banners linking to trusted sources. Beyond health crises, platforms also expanded governance mechanisms to address misinformation during election periods across countries such as South Africa [247], Ghana [248], Thailand [249], and the United States [250], as well as in response to security and violence-related events, such as the New Zealand terrorist

attack [251]. For example, in January 2024, Snap [250] outlined a comprehensive election integrity plan, including proactive removal of misinformation and deepfakes, human review and fact-checking of all political ads, bans on foreign-funded advertising, and real-time monitoring by its election integrity team to safeguard users from election-related harms. Similar efforts were seen in conflict-related contexts [252], where platforms prioritized the removal of harmful or misleading content to reduce societal risk.

Figure 3.7. Mechanisms of the digital wellbeing practices from company announcements



Note: Mechanisms refer to the specific actions, design features, and operational levers through which platforms implement their digital well-being practices, including both interface-level interventions and informational or governance measures. The mechanisms are derived from thematic analysis on company announcements. Data were collected from official news channels where social media platforms publish company announcements.

A further set of company announcements report the outcomes of content moderation or describe measures implemented within the moderation process to enhance transparency.

Several platforms have begun publicly reporting their moderation outcomes and process improvements. In February 2021, TikTok [253] enhanced transparency by publishing a detailed Transparency Report that disclosed the volume, speed, and rationale for content removals, appeals, and ad rejections. In January 2024, YouTube [254] introduced a new Channel Activity feature for live chat moderation, allowing streamers and moderators to view users' past moderation history, such as deleted messages, timeouts, or hides from the previous 30 days, to support more informed and transparent decisions. Likewise, in December 2024, TikTok [255] further strengthened its fact-checking system through partnerships with 20 independent fact-checking organizations worldwide to verify information, label unverified content, restrict it from For You feeds, and proactively remove misinformation that violates its policies.

Beyond these three major areas, a smaller set of content-level practices also emerged, including AI-generated and synthetic content management, content labeling, and content reporting. AI-generated and synthetic content management refers to platform policies and tools designed to detect, label, or regulate content created or altered by artificial intelligence. For example, in February 2020, Twitter [256] introduced a policy on synthetic and manipulated media, pledging to label or remove harmful fabricated content and provide contextual information to improve users' understanding of media authenticity. In April 2024, Meta [257] updated its policy to label AI-generated or manipulated media with "AI info" tags, prioritizing transparency and contextualization over removal and applying stricter labels to high-risk or deceptive cases.

Another content-level approach involves content labeling, both for promoting reliable information and warning against potentially misleading material. For instance, in May 2024, TikTok [258] launched a dedicated STEM feed in Europe, where all featured content is fact-checked and labeled by collaborating institutes to ensure accuracy and educational quality. In March 2022, TikTok [259] began labeling state-controlled media and adding warning labels to unverifiable or misleading videos to provide users with greater context and transparency.

Finally, several initiatives focus on content reporting systems, which empower users to flag harmful or unwanted material and enhance accountability in moderation. For instance, in April 2025, Snap [260] improved its content reporting mechanism by enabling users to easily report individual messages, stories, or accounts, including specific categories for non-consensual intimate imagery (NCII) and child sexual exploitation, allowing its 24/7 safety teams to prioritize and respond rapidly. Similarly, in February 2023, TikTok [261] introduced a “Report records” page that allows users to track the status of their submitted reports, thereby enhancing transparency and user trust in the review and enforcement process.

User-initiated experience control

Another category of mechanisms observed involves built-in platform controls and tools that users can actively configure, including parental controls, user management features for screen time, content, and contacts, as well as user management for compromised accounts.

Parental control refers to platform-provided tools and settings that enable parents or guardians to monitor, manage, and restrict their children’s online activities to ensure age-appropriate, safe, and balanced digital experiences. For example, In November 2020, TikTok [262] expanded its Family Pairing feature, allowing parents to link their accounts with their teens’ to manage safety and privacy settings, including controls over search, screen time, restricted mode, comments, account discoverability, liked videos visibility, and direct messaging. In February 2021, YouTube [263] introduced supervised accounts that allow parents to manage their tweens’ and teens’ access through three content settings, Explore, Explore More, and Most of YouTube, while controlling features like watch history, search history. Similarly, In March 2022, Meta launched Family Center and new parental supervision tools on Instagram [264], enabling parents to monitor teens’ screen time, manage interactions, set usage limits, and access expert educational resources, all within a centralized hub designed to support safe and age-appropriate digital experiences.

Another category concerns user management of screen time, comments, contacts, and content. Many of these features are optional for general users but enabled by default for minors under 18. For instance, in March 2023, TikTok [265] introduced a For You feed refresh feature that lets users reset content recommendations and added options to filter specific hashtags or mark videos as “not interested,” enabling greater control over viewing experiences. In May 2025, TikTok [266] further launched a “Sleep Timer” feature to promote mindful screen use by automatically interrupting teens’ feeds after 10 p.m. with guided meditation exercises, enabled by default for teens and optional for adults. Likewise, in June 2022, Meta [267] introduced the “Hidden Words” feature to help teens manage their online interactions and foster safer, more positive experiences on Facebook and Instagram. Similarly, YouTube [268] enhanced its screen time tools by making “Take a Break” and “Bedtime” reminders more prominent and frequent, with default hourly breaks for users under 18 and adjustable settings for all viewers. Platforms also expanded tools to help users understand and manage what content they see. In August 2023, Snapchat [269] introduced new controls, in compliance with the DSA, that allow EU users to see why content is recommended to them and opt out of personalized Discover and Spotlight feeds.

Beyond user control over screen time and content, platforms also provide visibility controls, which are tools allowing users to manage who can contact or interact with users. For example, in November 2021, Facebook [270] strengthened user control over visibility and interactions through tools such as unfriending, unfollowing, blocking, restricting, comment and tag controls, and Limits, which let users manage who can see, mention, or contact them and hide unwanted or abusive content on Facebook and Instagram. Similarly, in February 2025, TikTok [271] launched an in-app #SaferTogether hub that helps users manage visibility and interactions by adjusting account settings (who can discover or view their content), community controls (who can follow, message, or comment), and content controls (filtering keywords, marking videos as “Not Interested,” or enabling Restricted Mode). In September 2023, WhatsApp [272] introduced Channels, giving users greater control over their visibility by protecting personal information of followers and

admins, allowing admins to decide who can follow or discover their channel, and keeping users' follow choices private.

Social media platforms also implement digital well-being practices aimed at managing compromised user accounts, which occur when unauthorized parties gain access to a user's credentials. For instance, in April 2018, Meta [273] enhanced its account security infrastructure by disabling search-by-phone and search-by-email functions and by revising account-recovery procedures to mitigate data scraping and reduce the risk of account compromise. Similarly, in March 2025, TikTok [274] introduced Security Checkup, an in-app dashboard designed to help users detect and respond to potential account breaches by reviewing login devices, activating two-step verification, monitoring suspicious activity, and updating recovery options to strengthen overall account protection.

Default access and account protection

Another mechanism employed by social media platforms to promote digital well-being involves default access controls and account protection measures. These actions are typically implemented directly by the platforms and are not adjustable by users. Such measures include age restrictions, data security features, and account protection, verification, and removal protocols.

Age restriction refers to platform-enforced measures that limit access to specific features or content for users below a certain age, aiming to protect minors from harmful material and ensure age-appropriate online experiences. For example, in July 2021, Meta [275] stated that Facebook and Instagram are not designed for users under 13 and that the company enforces minimum age requirements through age screens, AI-based age detection, and account removal for underage users. In November 2020, TikTok [276] strengthened its safeguards by restricting Direct Messaging to users aged 16 and above to prevent potential misuse of the feature. Similarly, in April 2023, Snapchat [277] enhanced its age verification system by using users' birthdates to

ensure its My AI chatbot delivers age-appropriate responses and promotes safer interactions for younger audiences.

In addition, social media platforms claim to employ data security features, which are technical measures designed to safeguard user information from unauthorized access, misuse, or breaches. For instance, in December 2021, Meta [278] improved data security in Messenger and Instagram Direct Messages by making end-to-end encryption (E2EE) the default setting and advancing abuse detection mechanisms to protect user privacy while mitigating harmful behaviors. Likewise, in September 2023, WhatsApp [279] expanded its Channels feature globally, introducing privacy safeguards that hide follower identities, protect admin and user information, and automatically delete updates after 30 days. In April 2025, TikTok [280] strengthened data protection for European users by bringing its Norwegian data center online, enabling local data storage, independent oversight, and enhanced privacy controls.

Furthermore, social media platforms report improving user accounts security by enhancing account protection, improving verification, and strengthening removal of inauthentic accounts. Early efforts focused on account protection. For instance, as early as October 2011, YouTube [281] strengthened user account security by promoting the use of unique passwords, implementing two-step verification, and educating users about phishing, malware, and other online threats. Platforms have also developed in-app protection tools that automatically detect and restrict harmful content in direct messages. For example, in April 2024, Meta [282] introduced nudity protection in DMs, which uses on-device machine learning to blur images containing nudity, display cautionary messages to senders and recipients, while keeping image analysis local to maintain user privacy. In addition, platforms have introduced account verification measures. For example, in May 2020, Facebook [283] implemented an identity verification system for high-reach or potentially inauthentic profiles, requiring users to submit valid identification if their posts went viral or showed suspicious patterns, and those who refused or failed verification had their content distribution reduced.

Besides account protection and verification, platforms also engage in account removal to maintain user safety and platform integrity. For instance, in its 2019 report, Meta [284] stated that it handles fake accounts through automated detection systems that block them before creation, remove them immediately after signup, and disable those later found active. Similarly, in a 2018 post, X (formerly Twitter) [285] reported that it suspends accounts involved in harassment or manipulation by combining automated systems and human review to swiftly detect, remove, and block users who violate policies or attempt to disrupt public conversations during breaking events. In February 2023, TikTok [261] announced that it permanently bans accounts that repeatedly or severely violate its Community Guidelines, and further emphasized its zero-tolerance enforcement policies [286] in April 2024.

Personalized algorithmic content curation

The content users encounter on social media platforms is largely determined by personalized algorithmic recommendation systems. These algorithms operate by mapping the vast universe of potential content a user might engage with, identifying items most aligned with inferred preferences, and ultimately prioritizing and recommending those that best serve the platform's optimization goals [22]. Platform decisions regarding how to interpret user behavior signals and how to balance engagement, retention, and safety directly shape what appears in users' feeds. Yet these decisions are largely opaque: platforms rarely disclose the specific signals they prioritize, the weights assigned to engagement versus safety objectives, or how algorithmic trade-offs are resolved in practice. In response to growing concerns about the psychological and social effects of algorithmic personalization, platforms have introduced several digital well-being practices aimed at improving the recommendation process itself. These include adjusting feed composition, and limiting user data collection, and enhancing advertisement controls to promote user experience.

Several company announcements explicitly describe how their recommendation systems operate. A 2019 YouTube [287] post provides an overview of how the platform uses user data, including watch history, search history, clicks, watch time, likes, dislikes, shares, and survey-based satisfaction signals, to power personalized recommendations. Similarly, in a 2023 post, Meta [288] explains that its recommendation system identifies relevant content for users by (1) interpreting the semantic meaning of multimodal content through models such as MViT, XLM-R/XLM-V, and FLAVA/Omnivore, and (2) employing large-scale retrieval systems that rapidly narrow billions of candidate items to a few hundred most aligned with an individual user's interests. Meta [289] further updated the ranking and content information in 2025. However, these posts could be hard for non-technical people to understand. Furthermore, we do not observe comprehensive reporting on how these recommendation systems function, nor do we see long-term tracking of their development and performance over time.

Beyond offering high-level explanations of how their recommendation systems operate, platforms have also taken steps to increase transparency for users. For example, Meta [290] shared approaches in publishing “system cards,” disclosing key ranking signals, creating posts within transparency centers, and expanding in-app tools such as “Why Am I Seeing This?” Meta has further introduced user controls that allow people to personalize their feed experience by incorporating explicit user feedback: Instagram [291] has offered a “Not Interested” option since 2021, and in 2023 Meta added the “Show More / Show Less” feature [292], enabling users to directly signal whether they want to see more or less of specific types of content. However, these approaches also place the burden of algorithmic correction on users themselves, requiring active and ongoing effort from users.

Beyond announcements describing how their recommendation systems operate, platforms also claim to be adjusting the composition of recommended content, some by increasing diversity in users' feeds and others by reducing the visibility of borderline content. For example, in December 2021, TikTok [293] outlined several efforts to improve its recommendation algorithm

by diversifying For You feed content, limiting repetitive or sensitive videos, enabling users to filter unwanted topics, and embedding safeguards to ensure recommendations remain varied, personalized, and safe. More specifically, TikTok reported diversifying recommendations by interspersing content outside users' typical preferences, preventing consecutive videos from the same creator or sound, and testing mechanisms to avoid clusters of similar content (e.g., dieting, sadness, breakups). In 2023, TikTok [294] also introduced an option for EU users to disable personalization entirely to comply with the Digital Services Act (DSA). Snapchat proposed a slightly different approach. In a November 2020 post, Snapchat [295] emphasized rewarding creativity on Spotlight by distributing over \$1 million per day to eligible creators using a proprietary formula based on unique daily views, while monitoring for fraud and requiring compliance with Spotlight guidelines. In contrast, YouTube's 2022 priorities [296] highlighted efforts to reduce the prevalence of "borderline" content, videos that approach but do not violate policies, with a stated goal of keeping such content below 0.5% of total views on the platform.

Beyond platform-level decisions about recommendation system design, another important factor shaping what users see is the set of metrics platforms use to rank content, such as engagement rate, retention, or valued watch time. For example, Facebook's 2018 "time well spent" update [297] sought to reduce passive public content and prioritize meaningful interactions with friends and family, yet it did not specify which user metrics were used to operationalize or quantify the "time well spent" concept. More broadly, we found no company announcements that directly explain how these metrics are defined, weighted, or optimized within ranking algorithms. Instead of disclosing the metrics underlying engagement optimization, platforms address these issues indirectly, typically through user-facing controls that influence visible engagement signals without revealing how those signals are treated in the ranking process. For example, in May 2021, Instagram and Facebook [298] introduced options to hide like counts and filter comment interactions, tools that may influence how users engage with content but do not alter how engagement metrics are internally used for personalized recommendations. Similarly, YouTube

[296] removed public dislike counts to reduce harassment, yet continues to incorporate individual dislike signals internally to shape recommendation outcomes.

Another important dimension shaping the output of personalized recommendation systems is the set of user data that platforms track and convert into algorithmic signals. However, across platforms we found little detailed information about which specific user data streams ultimately become inputs to their personalization models. What platforms do describe instead relates primarily to restrictions on data usage for advertising, especially for teens. For instance, in 2023 Meta [299] announced that advertisers can no longer target teens based on prior engagement data, such as Instagram posts they have liked or Facebook Pages they follow, and that only age and location may be used for ad targeting, with gender removed as an option. Similarly, TikTok [300] explains in an August 2022 post that it uses data such as age, gender settings, and inferred interest categories to personalize ads, and that its “About this ad” feature allows users to see which data informed a specific ad and adjust their ad personalization preferences. Snapchat has made comparable changes. Since August 2023, to comply with the EU Digital Services Act, most targeting and optimization tools on Snapchat [301] are no longer used to personalize ads for certain users, and the platform now provides expanded transparency about why a particular ad was shown and gives users the ability to control their ad personalization settings.

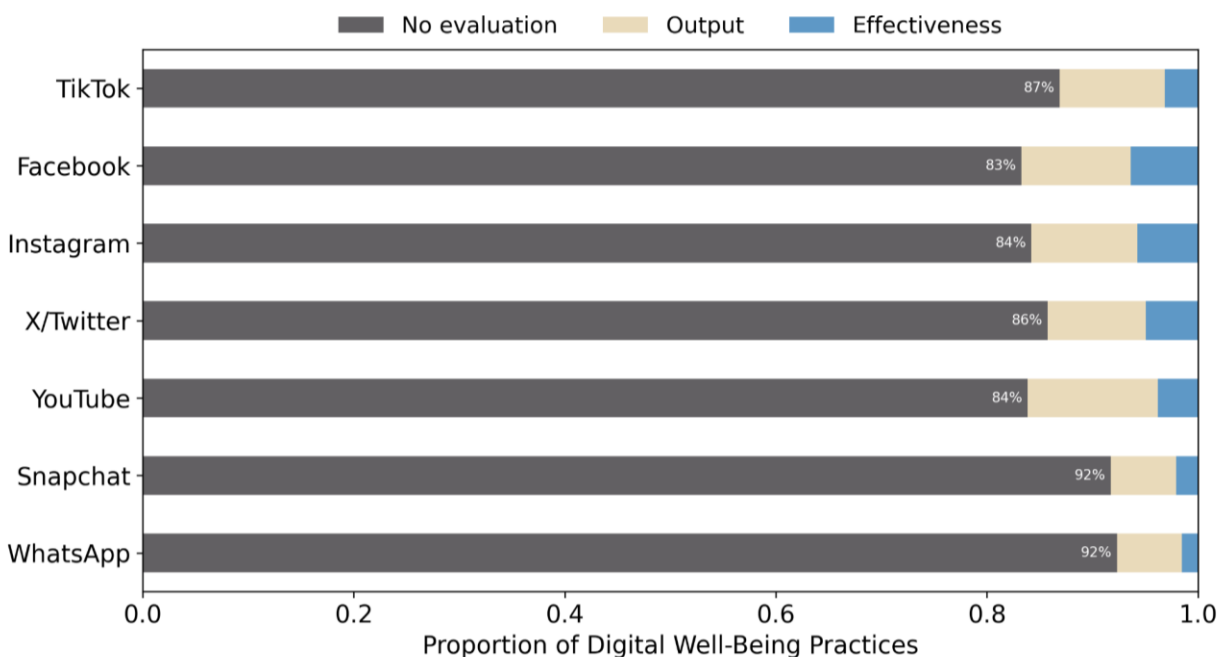
4.3.7 Impact evaluation status of digital well-being practices

In evaluating platform-initiated digital well-being practices, we observe that most company announcements describe the activities undertaken, whereas what is also important, but less frequently reported, are the outputs and effects of these activities. Outputs refer to the immediate, measurable products of an action. For example, the number of posts removed for violating platform policies, the number of accounts flagged or suspended, or the introduction of a new safety or screen-time feature. Effects capture the short- or long-term changes resulting from these

outputs, such as reductions in users' exposure to harmful content or improvements in users' self-reported well-being.

Across platforms, we found that most company announcements did not report either type of information. Shown in Figure 3.8, approximately more than 83% of company announcements across platforms did not include any evaluation details. Only about 5-10% of posts mentioned outputs, such as the number of violating posts removed. Fewer than 5% mentioned effects, such as changes in user safety outcomes or reductions in harmful content exposure. For example, in a TikTok post [302] introducing its mental health awareness collaboration with WHO, the company highlights activities such as partnering with WHO, engaging Fides creators, and providing mental-health resources, as well as inputs such as a \$3 million donation and access to expert networks, but it does not report any outputs or outcomes associated with these efforts.

Figure 3.8. Evaluation status of the digital wellbeing practices from company announcements



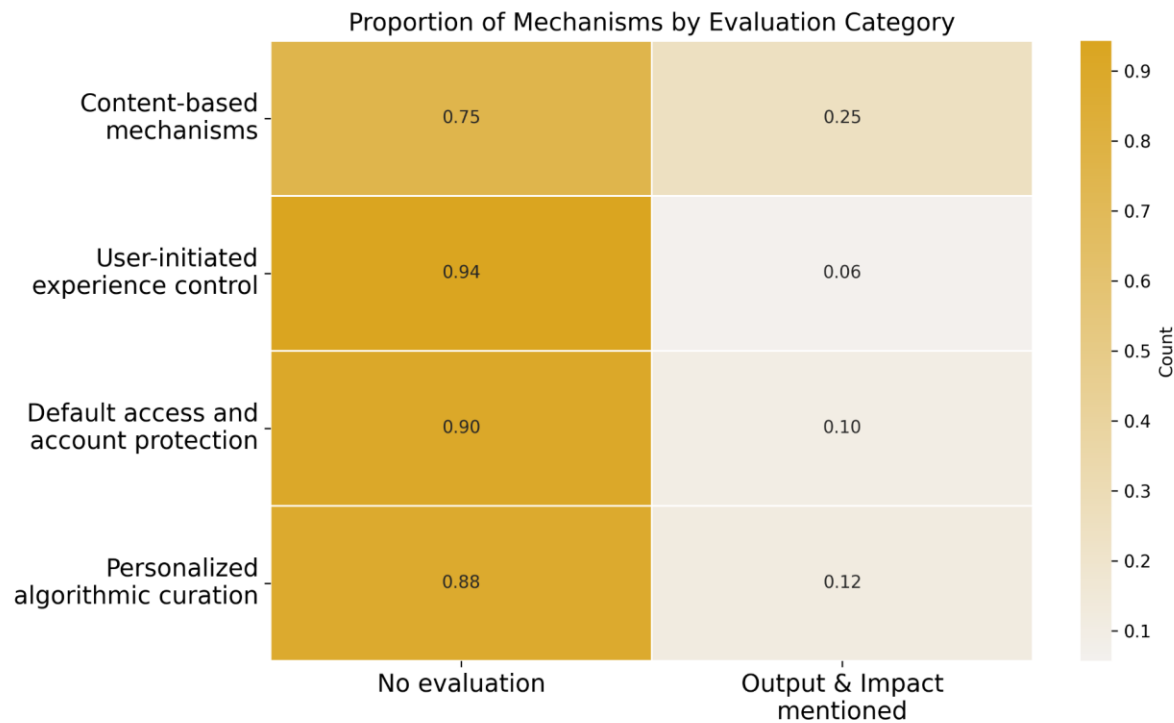
Note: The plot illustrates the primary evaluation status addressed by each platform's digital well-being practices. Data were collected from official news channels where social media platforms publish company announcements. Figure A3.8 presents the raw number of digital well-being practices categorized by platform and user harms.

When examining differences across the four mechanism categories, we found that content-based mechanisms contained the highest proportion of reported evaluation information (Figure 3.9). These posts most often cited outputs such as the volume of content removed or accounts closed. For example, in January 2021, YouTube [303] reported removing more than half a million videos under its COVID-19 policies since February 2020. Similarly, an August 2020 TikTok post [304] on countering hate speech reported removing 380,000 videos, banning 1,300 accounts, and deleting 64,000 hateful comments; however, the post acknowledged that these figures do not represent a comprehensive capture of all hateful content, and it did not provide more effect-level evaluations, such as reductions in users' exposure to hate or improvements in user self-reported perceived safety. A December 2022 post from Meta [305] likewise mentioned blocking millions of fake accounts per day and removing more than 150 networks of malicious actors since 2017, and hate speech exposure on Facebook decreased from 10–11 to about 2 views per 10,000 content views. In general, effect evaluations for the platform practices through content-based mechanisms remain rare, and virtually none draw on user self-report data, even for short-term outcomes.

In contrast, posts related to user-initiated controls, default access and account protections, and personalized algorithms exhibit a much lower prevalence of reported output or effect evaluation, with only a few exceptions. For user-initiated controls and for data-safety or default-access protections, platforms typically report outputs such as the frequency of feature use or the number of accounts reviewed or removed, but provide little evidence of downstream effects. For example, in a post announcing the expansion of its Mental Health Education Fund and the launch of guided meditation features, TikTok [306] reported several outputs, including 173 million impressions, 600,000 new followers, 200,000 website visits, 486 new volunteers to the mental health awareness topics, and a 98% retention rate for the use of the Sleep Schedule feature. Likewise, a Meta post [278] about safer private messaging shared outputs such as more than 100 million views of safety notices in one month and a 50% year-over-year increase in reporting.

Similarly, a June 2025 Snapchat [307] post described a pop-up warning for teens that resulted in more than 12 million blocks, yet provided no evidence of subsequent effects on user safety or experience.

Figure 3.9. Evaluation status of the digital wellbeing practices from company announcements



Note: The plot illustrates the primary evaluation status addressed by each platform’s digital well-being practices. Data were collected from official news channels where social media platforms publish company announcements.

Impact evaluation is even more limited for personalized algorithm mechanisms, as platforms rarely explain how algorithmic changes translate into measurable outputs or outcomes. For example, Meta [290,308] evaluated its recommendation system by asserting that recent AI advancements contributed to a 15% increase in watch time in the Facebook Reels player, along with reduced skip and hide rates. However, the post did not provide sufficient detail on how these metrics were generated or assessed, limiting the interpretability of these claims for external readers.

4.4 Discussion

4.4.1 *Summary of findings*

In this study, we conducted a large language model assisted content analysis of all 13,497 company announcements published by seven major social media platforms from their inception through May 15, 2025. We identified 5,685 digital well-being practices and described their distribution by platform, target user groups, harm categories, mechanisms, and impact evaluation status. Our findings show that platforms have reported an increasing number of digital well-being practices in their public announcements since 2019, with a notable rise following the introduction of the EU General Data Protection Regulation in 2018. We also found that TikTok's country and regional offices in the Global North, particularly Germany, the United States, and the EU regional office, reported more digital well-being practices than offices in the Global South. Approximately one third of all practices focused on addressing harmful content, such as hate speech and offensive material, followed by practices aimed at reducing harmful contact to protect users from abuse, harassment, and bullying. By contrast, relatively few practices addressed harmful behaviors such as excessive scrolling.

We classified the digital well-being practices into four broad mechanisms: content-based mechanisms, user-initiated experience controls, default data safety and account protection, and personalized algorithmic content curation. Most practices fell within content-based governance or user-initiated controls, while very few involved changes to personalized algorithmic curation. Across all four mechanisms, platforms provided limited evidence of actual user level impact. Announcements rarely reported how many users were affected by a practice, how users experienced the change, or whether the practice produced the intended effects. Content based practices sometimes reported the number of posts or accounts removed, yet these metrics remained disconnected from user level outcomes. Only a small number of announcements mentioned user self-reported experiences, and platforms disclosed even less about the impacts of privacy measures, safety features, or algorithmic adjustments.

TikTok exhibits a comparatively higher volume of digital well-being announcements in our analysis. However, we cannot determine whether this pattern reflects greater substantive investment in digital well-being initiatives or a more intensive signaling strategy. Notably, many of TikTok’s announcements are concentrated in content governance and health-related information. This emphasis may reflect stronger incentives to publicly demonstrate responsibility, particularly given TikTok’s heightened geopolitical scrutiny and regulatory vulnerability as a non–U.S.-origin platform.

Taken together, these patterns suggest that social media platforms increasingly signal a commitment to digital well-being, particularly after 2019 amid rising public concern and regulatory pressure. Yet the practices they report focus mainly on content removal, information provision, and optional user-initiated controls, with far fewer efforts aimed at improving default data protections or algorithmic systems. Moreover, platforms offer little evidence about whether these actions improve users’ experiences. As a result, the actual implementation and the effectiveness of platform driven digital well-being practices remains largely unknown.

4.4.2 Transparency is needed in reporting what platforms are actually doing

A central challenge in understanding platforms’ digital well-being efforts is the persistent gap between what companies claim in their company announcements and what they actually implement. This gap arises for three main reasons: (1) platform announcements may not accurately describe how features work in practice, (2) the target populations and geographic coverage of digital well-being practices are often unclear, and (3) platform designs change frequently, making it difficult to track whether a feature is new, updated, or replaced.

First, public descriptions of digital well-being features do not always match their actual operation. For instance, according to the *North Carolina v. TikTok* complaint, TikTok’s widely promoted 60-minute daily limit for teens does not function as an actual limit. Instead, teen users merely receive a notification after 60 minutes and can continue using the platform by entering a

simple preset passcode. This illustrates how company narratives may overstate or misrepresent the protective value of safety features. Second, platform announcements rarely specify where or to whom a feature applies. Some digital well-being tools are launched only in specific countries due to local regulations, such as features introduced in the European Union to comply with the Digital Services Act. It is unclear whether these features are also applied in other regions. Third, digital well-being features evolve constantly, and the current company-announcement format provides only a fragmented and incomplete record. For example, a single feature may appear multiple times in announcements or be described differently across updates, making it unclear whether it represents a revision of an existing feature or an entirely new one. Although we attempted to identify duplicative features, this effort is constrained by the absence of a consistent system for documenting updates or tracking regional differences.

In summary, these factors make it nearly impossible for external observers to verify what platforms have truly implemented, how broadly a practice is applied, let alone what effects it produces. This persistent gap between public claims and internal operations underscores the need for a more transparent and systematic disclosure system for platform digital well-being practices.

4.4.3 Lack of transparency in the impact of digital well-being practices

Beyond the challenge of understanding what platforms actually implement, another major challenge concerns the unclear effects of digital well-being practices on users. We found that the majority of company announcements provide no information about the impact of these practices, raising concerns that the digital well-being features described may function as cosmetic practices or symbolic signals of responsibility than as effective interventions.

For content-based mechanisms, some announcements report the number of posts, accounts, or pieces of content removed. Similarly, companies such as Meta and TikTok have begun releasing transparency reports [309] that highlight metrics such as the volume of content

taken down. Yet the definitions behind these metrics are not clearly articulated. More importantly, aggregate removal counts cannot indicate how many users actually encountered harmful content or how many user sessions were protected. Given that social media platforms host billions of posts each day, it is difficult to map moderation numbers onto the actual number of users affected. A reduction to one percent of harmful content may sound encouraging, but at global scale, one percent still represents substantial exposure at the user level. Moreover, the effectiveness of the content moderation model itself has been challenged. A study [310] found content removals prevented only 24–30% of the posts’ predicted engagement. For many announcements aimed at providing well-being-related information or instructions for using certain features, it also remains unclear whether users see or access these materials. In short, without information linking content-based mechanisms to user experience, it is impossible to know what these metrics mean for real users.

For user-initiated controls, our study found that only a small share of announcements report any outcomes. External studies [199,217,219] consistently document low compliance and high attrition with self-control tools for digital well-being, such as screen time reminders. If the compliance and effectiveness of external tools are low, it is reasonable to question whether internal platform features requiring similar self-regulation are widely adopted or meaningfully affect user experience. Platforms may highlight the availability of these controls, but we do not know how many users use them, how often they use them, or whether they lead to changes in screen time or well-being.

The effects of algorithmic recommendation systems are even less transparent. Platforms occasionally make public claims about improving feed ranking, such as Facebook’s 2018 announcement that it would prioritize meaningful social interactions [311]. However, evidence about the impact of such changes is limited, and many initiatives have since faded from public discussion. For external researchers, personalized algorithms are complex systems, and quantifying the effects of digital well-being practices on algorithmic recommendations is even

more challenging. One of the closest studies, a randomized controlled trial by Guess et al. [312], found that turning off personalized recommendations and switching to reverse chronological feeds on Facebook and Instagram reduced screen time, but the amount of political and untrustworthy content users saw increased. Recent research [22] indicates that personalized algorithms do not need to be an all-or-nothing choice between default personalization and chronological feeds. Users can be given meaningful ways to adjust their recommendations, such as selecting among different recommender systems or setting preference filters, and platforms could set defaults optimized for long-term well-being. However, none of these user-directed or value-driven approaches appear in the company announcements we reviewed.

4.4.4 Policy implication: global governance with evidence-based and transparent infrastructure

Our findings suggest that company newsroom announcements are not a sufficient mechanism for transparency regarding digital well-being practices. To support meaningful oversight, the global community will need more robust systems for documenting platform design choices and for reporting their effects on users in a consistent and accessible way. Based on this study, we propose three features that an ideal transparency system for platform digital well-being design should include: (1) a clear harm-responsive design process, (2) collaboration with external researchers or evaluators to assess impact in a manner analogous to clinical trials for new medicines, and (3) a global governance framework to ensure equitable access and accountability.

First, an ideal transparency system should follow a harm-responsive design process, clearly articulating why platforms introduce specific digital well-being features, how these features are developed, and what effects they produce. A complete system would mirror core steps in academic research: platforms would describe the problem they aim to address, explain why the problem matters based on internal evidence or monitoring, outline the actions they plan to take, and disclose both the outputs and the short- and long-term effects of those actions on user

behavior, well-being, and mental health. Many platforms already employ internal research teams capable of generating such information. A dedicated disclosure structure that shares key elements of this process with the public would make digital well-being practices far more understandable and accountable.

Second, the process could draw from the structure of clinical trials for new medicines, where external researchers and evaluators have independent access to data and content to assess impact. When a pharmaceutical company develops a medical treatment, it must follow a clear protocol with multiple phases of testing, each with defined goals, participants, and evaluation criteria. External experts are involved at specific stages, and regulatory bodies such as the Food and Drug Administration (FDA) oversee the process through established standards. Social media platforms also produce products that affect behavior, well-being, and health. Granting external researchers free and open access to platform data and content would enable independent audits of platform practices and their effects, reducing reliance on self-reported metrics. Although platform products are more complex and cannot be isolated as easily as a drug compound, the underlying principle remains relevant. A structured system could require platforms to explain what a new design feature aims to do, who it might affect, when external reviewers should evaluate it, and which institutions at national or global levels should provide oversight. Such a protocol would give the public a clearer understanding of platform decision-making and establish a consistent pathway for assessing safety and effectiveness.

Finally, a complete transparency and safety system will require global governance. Other public spheres, such as health and environmental protection, rely on international bodies that set shared standards and coordinate action, including the WHO and United Nation Environment Programme. These systems are imperfect, but they reduce inequality between regions and prevent countries from addressing complex problems in isolation. Digital platforms operate globally, yet no comparable structure exists for transparency, safety, or accountability. Without a coordinated international framework, platform governance will continue to vary widely across

regions, reproducing inequities observed in public health and environmental governance. Developing a global mechanism for digital platform oversight would promote consistent standards and support a more equitable approach to user safety worldwide.

Building such systems is critical not only for social media governance but also for the future of AI. The shortcomings observed in social media safety, including unclear reporting, weak public oversight, and optimizing for maximum engagement but not long-term user value, demonstrate how difficult it is to steer complex digital systems toward user well-being once they are widespread. These lessons apply directly to AI products, from chatbots to emerging AI agents. The current moment provides an opportunity to design governance systems that center human values. Work on human-centered and value-aligned AI [313–315], including frameworks on human flourishing and purpose-driven AI, highlights the importance of aligning digital products with user goals rather than engagement metrics or profit incentives. Applying these lessons now can help ensure that AI systems evolve in ways that support user well-being rather than repeating the failures of social media governance.

4.4.5 Limitations

This study has several limitations. First, our analysis is based solely on what platforms choose to disclose through their public announcements. These announcements provide only a selective view of platform practices rather than a complete account of internal operations. It is possible that platforms report actions that do not reflect actual implementation. It is also possible that they engage in additional digital well-being work that is not reported because it is difficult to communicate to the public, particularly in areas such as data security, safety protocols, or algorithmic design. Because the study relies exclusively on publicly available newsroom announcements, it cannot validate the accuracy or completeness of what platforms claim. Other sources, such as the Facebook internal document archive released by whistleblowers or legal disclosures produced during lawsuits—including recent cases in which Meta was required to

release hundreds of pages of internal materials—offer partial insight into internal decision-making. However, these sources are limited, uneven, and often dated. As a result, our findings capture only the public-facing account of digital well-being practices.

Second, our classification of digital well-being practices relies on large language model assisted methods. Although we evaluated model performance and carried out additional checks to reduce misclassification, misclassifications are inevitable. Both false positives and false negatives may occur, and some practices may have been missed or incorrectly categorized. Additionally, LLM-assisted classification raises reproducibility concerns: outputs may vary across model versions, temperature settings, or prompt formulations, making exact replication by independent researchers difficult. To mitigate this, we used a fixed temperature setting, majority voting across repeated runs, and report our prompt structures in the supplementary materials, though full reproducibility cannot be guaranteed as model versions are updated over time.

Third, despite our efforts to remove duplicate entries, some repeated or overlapping announcements may remain in the dataset. Platforms sometimes publish the same announcement across different country or regional pages, especially in the case of TikTok. We attempted to detect duplicates by comparing URLs, comparing text directly, and using cosine similarity measures from sentence embeddings. When announcements were highly similar, we treated them as duplicates. Even with these steps, some announcements that describe the same feature in slightly different ways may have been counted more than once.

Fourth, there is no agreed upon taxonomy for platform design or digital well-being practices. Although we drew on prior work, including the platform design framework by Iyer and colleagues, existing classifications do not fully capture the range of practices we identified. As a result, our four mechanism categories were developed through a bottom up approach. This required researcher judgment, and other scholars may propose alternative ways of organizing these practices. We acknowledge that this is a source of subjectivity and welcome further discussion on how to build more comprehensive and standardized frameworks.

Finally, the digital environment is changing rapidly. By the time this study is published, policy shifts, platform design changes, or new governance actions may have already altered the landscape. For example, Australia's emerging policy to restrict social media access for users under sixteen may reshape both user behavior and platform responses. As platforms and regulators continue to evolve, the content of newsroom announcements and the nature of digital well-being practices may shift quickly. Our findings reflect the practices described during the period we studied but may not capture changes that occur afterward.

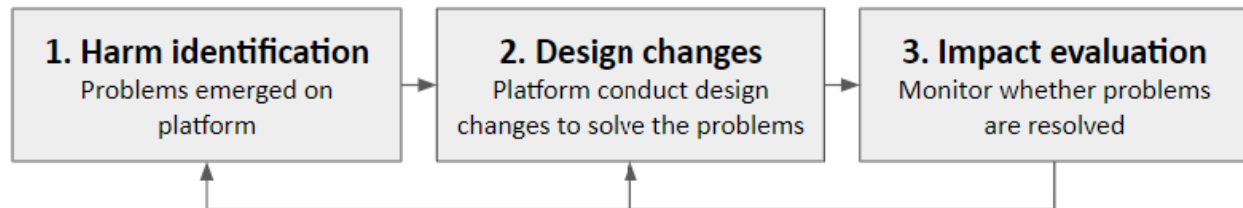
4.5 Conclusion

In conclusion, this study provides the first large-scale, LLM-assisted analysis of platform-reported digital well-being practices. By identifying 5,685 digital well-being practices across seven major social media platforms, we reveal clear trends: a notable expansion of platform attention to digital well-being since 2019, a greater volume of practices originating from Global North offices, and a particular focus on adolescents and youth. Although platforms increasingly signal a commitment to digital well-being, primarily through content removal, information provision, and optional user controls, they provide little evidence about how these actions affect users and rarely report efforts to strengthen default protections or algorithmic systems. These patterns highlight the limitations of company newsroom announcements as a transparency mechanism and underscore the need for more rigorous, evidence-based, and globally coordinated oversight. We argue that an effective transparency infrastructure for platform digital well-being practices should follow a harm-responsive design process, incorporate independent evaluation similar to clinical trial protocols, and be embedded within a global governance framework that ensures consistent standards and equitable protections for users worldwide.

Supplementary materials

Figures

Figure A3.1. Harm-responsive design process (HRDP) for assessing social media platform digital well-being design



Note:

- Harm identification: platforms assess harms and the design choices that enable them;
- Design changes: interventions are implemented to mitigate these harms;
- Impact evaluation: the effectiveness of these interventions, including outputs and outcomes, is systematically assessed.

HRDP provides a qualitative framework to determine whether platforms effectively identify harms, express their motivations for the design implementation, implement targeted or informed interventions, and conduct meaningful post-implementation evaluations.

Figure A3.2. Framework for LLM-Assisted Content Analysis Pipeline in Chapter 4

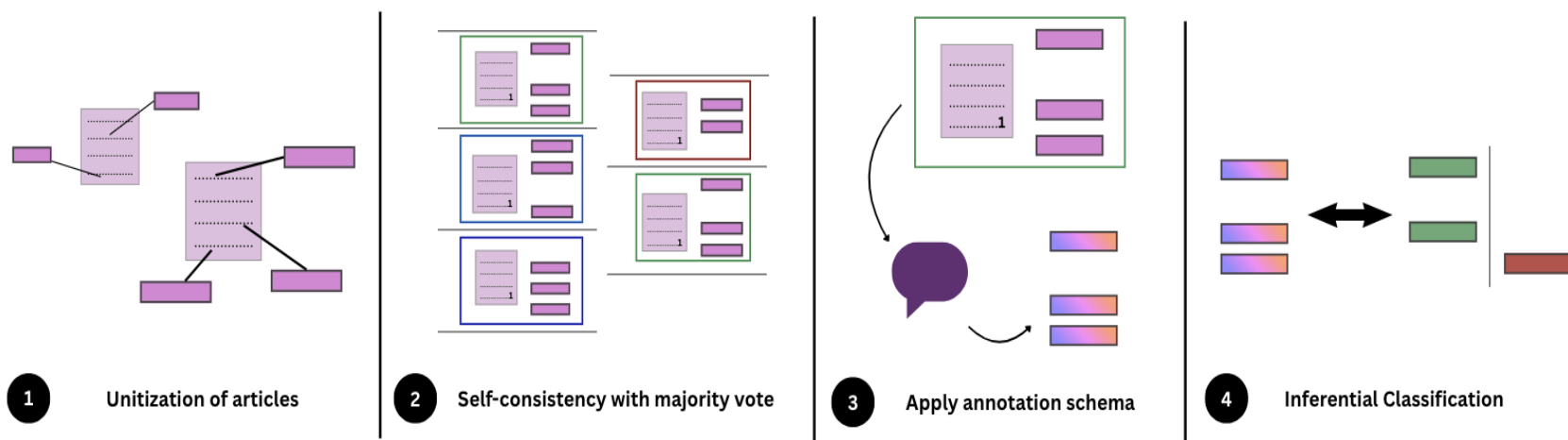


Figure A3.3. LLM-assisted pipeline for deductive content classification in Chapter 4

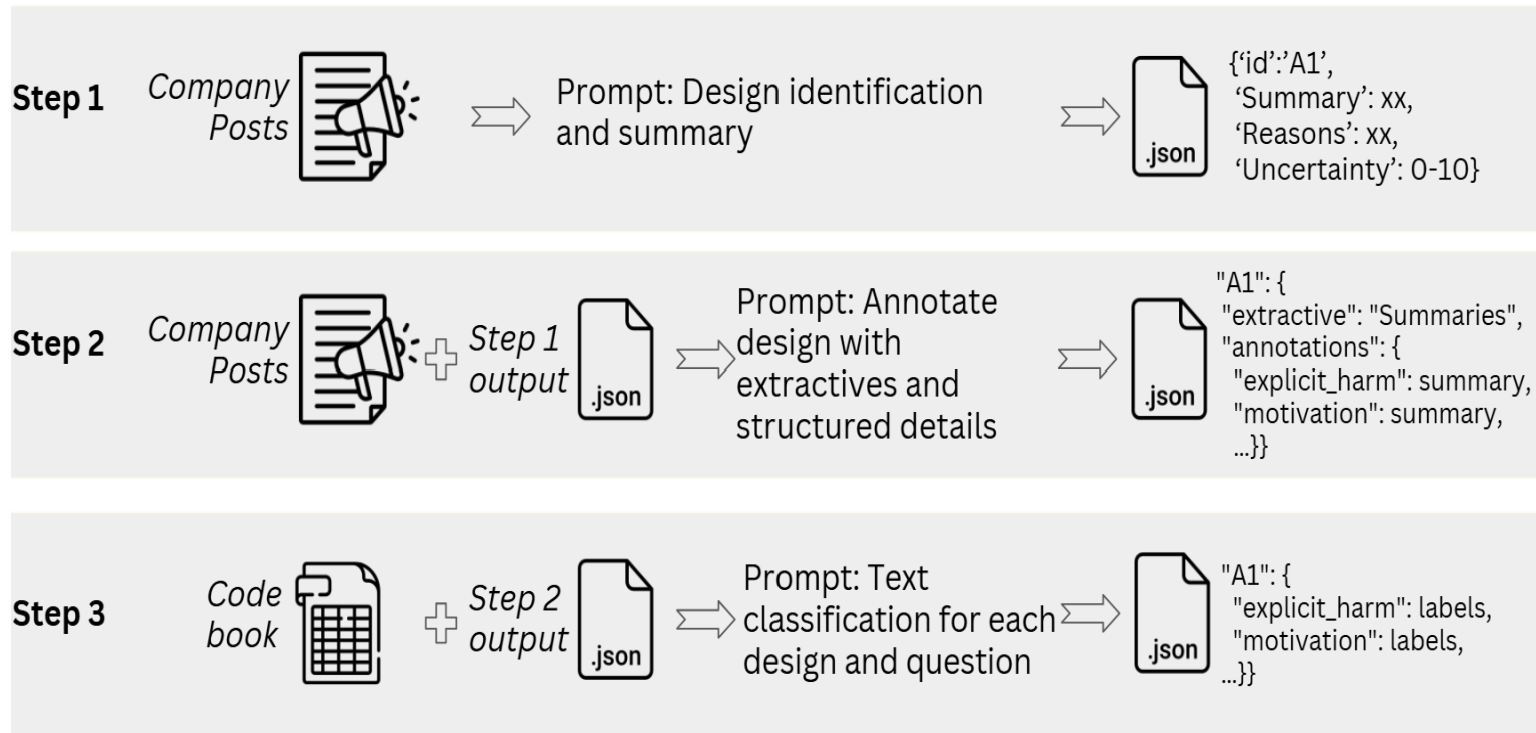


Figure A3.4. Example template of model response for Step 1

```
```json
{
 { "A1": Brief description of digital wellbeing design intervention,
 "abstractive": A brief summarization of the design change,
 "reason": "It satisfies inclusion criteria I1"
 },
 { "A2": Brief description of digital wellbeing design intervention,
 "abstractive": A brief summarization of the design change,
 "reason": "It satisfies inclusion criteria I1"
 },
 count: N
}
```

**Figure A3.5. Annotation Schema**

```

annotation_schema= {
 'explicit_harm': ""According to the **explicit_harm summary**, which harm category best describe the type of harm the PDWP is intended to address?""",
 'demographics': ""According to the **demographics summary**, which demographics best describe the type of target users of the PDWP?""",
 'motivation': ""According to the **motivation summary**, which type of motivation best explain why the platform introduced the PDWP?""",
 'motivation_evidence': ""According to the **motivation_evidence summary**, what was provided about what motivate the platform to initiate the PDWP (if any)?""",
 'motivation_nature': ""According to the **motivation_nature summary**, what type was provided about what motivate the platform to initiate the PDWP (if any)?""",
 'geographic_scope': ""According to the **geographic_scope summary**, what geographic PDWP apply to? Please list full names.""",
 'instruction': ""According to the **instruction summary**, what type of instruction is provided by the PDWP (if any)?""",
 'evaluation': ""According to the **evaluation summary**, what type of evaluation was mentioned in the PDWP?""",
 'evaluation_nature': ""According to the **evaluation_nature summary**, what type was provided about the PDWP (if any)?""",
 'partnership': ""According to the **partnership summary**, what types of partnership were mentioned in carrying out the PDWP?""",
 'entities': ""Which entity category (or categories) from the **annotation codebook entity (or entities) mentioned in the **entities summary**?""",
 'automaticity': ""Does it indicate that the change is: 1) applied by default and require user action, or 2) Requires users to apply, initiate, or access it?""",
 'design_element': "Which design element(s) from the list are referenced or implied?"
}

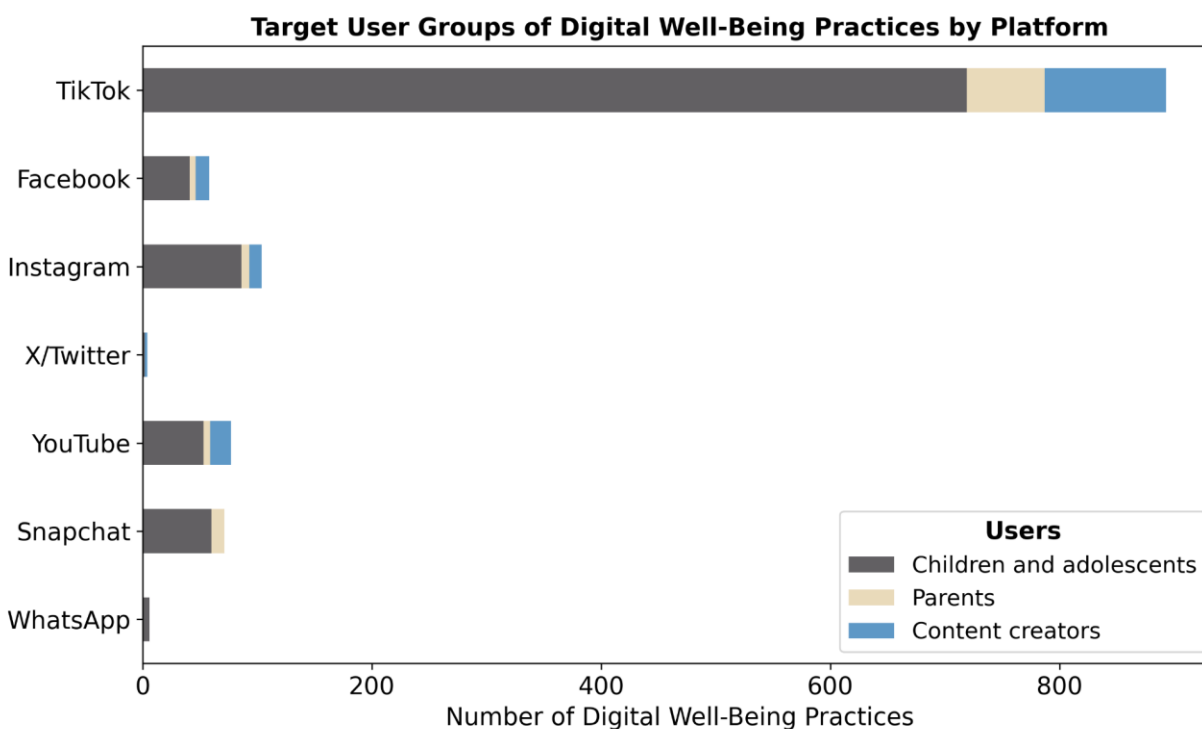
```

```

{
 "A1": {
 "extractive": Verbatim sentence(s) from the news announcement
 that you extracted for the specific PDWP",
 "annotations": {
 "explicit_harm": string (only if the news announcement mentions it)",
 "entities": string (only if the news announcement mentions it)",
 "motivation": string (only if the news announcement mentions it)",
 "motivation_evidence": string (only if the news announcement mentions it)",
 "partnership": string (only if the news announcement mentions it)",
 "geographic_scope": string (only if the news announcement mentions it)",
 "demographics": string (only if the news announcement mentions it)",
 "evaluation": string (only if the news announcement mentions it)",
 "instruction": string (only if the news announcement mentions it)"
 }
 }
}

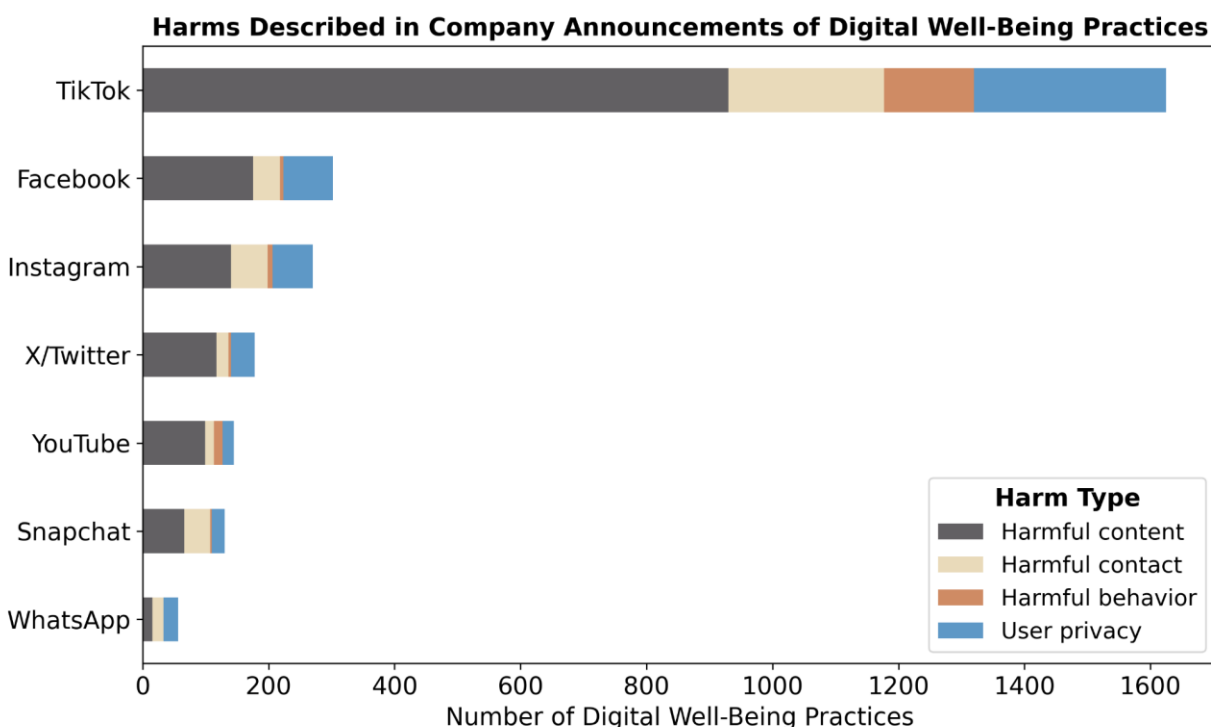
```

**Figure A3.6. Raw Counts of Digital Well-Being Practices by Platform and Target User**



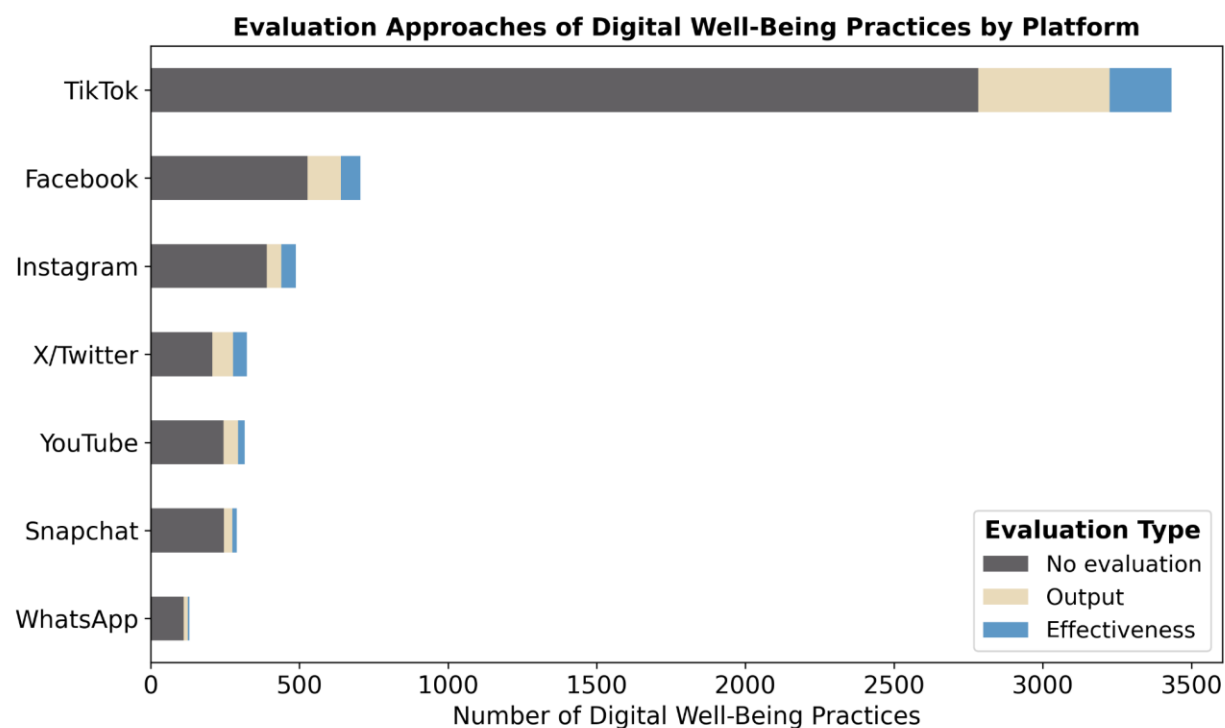
*Note: The chart illustrates the primary target users of each platform's digital well-being practices, that is, the intended audience specified in the company announcements. When the target audience was not explicitly stated, we assumed the practice was directed toward general users (not shown here). Practices explicitly target children and adolescents, parents, or content creators are displayed. This figure presents the raw number of digital well-being practices categorized by platform and target user.*

**Figure A3.7. Raw Counts of Digital Well-Being Practices by Platform and User Harms**



*Note: The plot illustrates the primary target harms addressed by each platform's digital well-being practices, categorized into four types based on the social media harm taxonomy: harmful content, referring to exposure to damaging materials such as misinformation, violence, nudity, or extreme content; harmful behavior, denoting negative psychological or behavioral reactions such as fear of missing out, social comparison, or loss of self-control; harmful contact, involving harmful interactions from other users or bots, including bullying or unwanted sexual advances; and harmful use of information (user privacy), concerning risks to user data and accounts, such as privacy breaches, impersonation, or malware exposure. Some digital well-being practices do not fall into any of these categories. For instance, those focused on providing health or positive information which are not the scope of this paper. The figure presents the raw number of digital well-being practices categorized by target harms.*

**Figure A3.8. Raw Counts of Digital Well-Being Practices by Platform and Evaluation Status**



*Note: The plot illustrates the primary evaluation status addressed by each platform's digital well-being practices. Data were collected from official news channels where social media platforms publish company announcements. This figure presents the raw number of digital well-being practices categorized by platform and user harms.*

## ***Tables***

***Table A3.1. The source of company announcements by social media platforms***

<b>Platform</b>	<b>Newsroom source</b>
Meta (Facebook, Instagram, WhatsApp)	<a href="https://about.fb.com/news/">https://about.fb.com/news/</a>
TikTok	<a href="https://newsroom.tiktok.com/">https://newsroom.tiktok.com/</a>
YouTube	<a href="https://blog.google/products/youtube/">https://blog.google/products/youtube/</a>
Snapchat	<a href="https://newsroom.snap.com/">https://newsroom.snap.com/</a>
Twitter / X	<a href="https://blog.twitter.com/">https://blog.twitter.com/</a>

**Table A3.2. Well-being-related tags and keywords used to filter digital well-being company announcements in the current study**

Inclusion filters	Tags / Keywords
Tags related to Well-being	Well-Being, Suicide Prevention, Security News, Safety, Privacy, Security, Health, Digital Literacy, Misinformation, Hate Speech, COVID-19 Response
Keywords related to Well-being	safety, safe, safer, secure, security, cybersecurity, privacy, transparency, transparent, trust, protection, enforcement, integrity, policy, policies, guidelines, parents, parent, child, children, young, youth, teenage, kid, student, well-being, wellbeing, experience, experiences, health, mental, prevention, fake, support, feature, awareness, responsible, sexual, rules, rule, school, bully, violence, disinformation, misinformation, strives, guardian, protect, harassment, authenticity, accountability, violation, suicide, suicidal

**Table A3.3. The performance of human and machine coder in the content analyses**

<b>Task in Content Analysis</b>	<b>Outcome Label (binary)</b>	<b>Agreement (between human coders)</b>	<b>Agreement (Human and machine coder)</b>
Include or exclude company posts		Kappa* = 0.71	Kappa = 0.69
Identify digital well-being practices in the posts		Kappa = 0.69	Kappa = 0.68
Target user groups	Children and adolescents	Kappa = 0.79	Kappa = 0.76
	Parents	Kappa = 0.78	Kappa = 0.76
	Content creators	Kappa = 0.78	Kappa = 0.75
Harms addressed	Harmful content	Kappa = 0.74	Kappa = 0.73
	Harmful contact	Kappa = 0.73	Kappa = 0.71
	Harmful behavior	Kappa = 0.76	Kappa = 0.72
	User privacy	Kappa = 0.75	Kappa = 0.70
Stated motivation	Company values	Kappa = 0.71	Kappa = 0.68
	Legislation	Kappa = 0.78	Kappa = 0.70
	External event	Kappa = 0.79	Kappa = 0.69
	External research	Kappa = 0.80	Kappa = 0.68
Evaluation process	No evaluation	Kappa = 0.82	Kappa = 0.75
	Output evaluation	Kappa = 0.75	Kappa = 0.68
	Effectiveness evaluation	Kappa = 0.78	Kappa = 0.67

## Codebook

---

### Metadata

#### URL

Provide the source URL of the news release or announcement.

---

#### ID

Enter the assigned ID for this item (e.g., A2).

---

---

### 1. Digital Well-Being Practice Presence

**Label:** *Digital Well-Being Mention*

**Question:** Does the extractive reflect a digital well-being practice mentioned on the URL page?

☐ Yes

☐ No (*If “No,” skip remaining items and submit response.*)

---

### 2. Explicit Harm Type

**Label:** *Explicit Harm Addressed*

Select all harms explicitly mentioned or targeted by the practice.

☐ Harmful content

☐ Harmful contact

☐ Harmful behavior

☐ Privacy-related concerns

☐ External harm (e.g., COVID-19)

☐ Transparency and accountability

---

### 3. Harm Label

**Label:** *Specific Harm Category*

Select all that apply (categories derived from thematic analysis).

☐ Not covered by any options below

☐ COVID-19

☐ Cancer

☐ Social Comparison

☐ General Mental Health Support

- ☐ LGBTQ+ Health and Safety
- ☐ Abuse, Harassment, and Bullying Prevention
- ☐ Body Image and Eating Disorders
- ☐ Suicide and Self-Harm
- ☐ Social and Emotional Learning
- ☐ Hate Speech and Offensive Content
- ☐ Terrorism and Violent Extremism
- ☐ Sexual Exploitation and Assault
- ☐ Accessibility for Disabled Users
- ☐ Digital Literacy
- ☐ Substance Abuse
- ☐ Sleep
- ☐ Anti-Racism and Racial Equity

---

#### 4. Motivation

**Label:** *Motivating Factors*

Select all sources of motivation cited for the practice.

- ☐ Company values
- ☐ Research
- ☐ Others
- ☐ Advocacy groups
- ☐ Community pressure
- ☐ Civil society
- ☐ Legislation
- ☐ Events

☐ Education

---

5. Motivation Nature

**Label:** *Nature of Motivation Citation*

If a motivation is cited, indicate its nature.

☐ Qualitative

☐ Quantitative—experimental

☐ Quantitative—observational

☐ Other

☐ Not mentioned

---

6. Automaticity

**Label:** *Implementation Type*

☐ By default

☐ User-initiated

---

7. Design Element (Platform Design Taxonomy)

**Label:** *Design Element (Set 1)*

Select all design elements mentioned.

☐ Not covered by any options below

☐ Infinite Scroll

☐ Auto Play

☐ Engagement Gamification

☐ Notifications (Platform or User Content)

☐ Requiring Account Creation

☐ Ephemeral Content

☐ Barriers to Deletion

☐ Attention Optimization

☐ Personalization

- ☐ Lack of Rate Limits
- ☐ Lack of Negative Feedback (Barriers to Reporting)
- ☐ Badges and Verification
- ☐ Discriminatory Algorithms
- ☐ Profile/Picture Effects
- ☐ Generative AI Platform Tools
- ☐ Generative AI User Tools
- ☐ Lack of Friend/Connection/Messaging Rate Limits
- ☐ Default Profile Privacy
- ☐ Default Post/Comment Privacy
- ☐ User Discoverability (Search or Friending)
- ☐ Age Verification
- ☐ Geolocation
- ☐ Disproportionate Data Collection by Platform

---

#### 8. Design Element (Thematic Coding)

**Label:** *Design Element (Set 2)*

Select all applicable elements.

- ☐ Not covered by any options below
- ☐ AI, Synthetic, and Manipulated Content Management
- ☐ Comment and Chat Message Moderation
- ☐ Educational and Policy Resources Provision
- ☐ Content Accessibility for Users
- ☐ Health and Well-being Information Provision
- ☐ Parental Controls

- ☐ Recommendation System Content Interventions
- ☐ Screen Time Management Tools
- ☐ Content Quality Labels
- ☐ Content Age Restriction Warning Labels
- ☐ Interaction Visibility Control
- ☐ Issue-Based Content Governance
- ☐ Data Security Features
- ☐ Compromised Account Detection
- ☐ Advertising and Monetization Controls
- ☐ Content Reporting and Appeals
- ☐ Account Verification Measures
- ☐ Account Protection Measures
- ☐ Content Moderation
- ☐ Misinformation Detection and Reduction
- ☐ User Content Control
- ☐ Data Collection Limits for Vulnerable Users

---

## 9. Entities Mentioned

### **Label:** *Entities Referenced*

Indicate any entity mentioned in the release.

*(Note: Partnership is a subset of entity—list all entities even if collaboration is implied.)*

- ☐ Businesses
- ☐ Non-academic research institutions
- ☐ Nonprofits and community groups
- ☐ Others
- ☐ Governmental agencies

☐ Academic institutions

---

#### 10. Partnerships

**Label:** *Collaborative Partners*

Select all entities explicitly described as partners.

☐ Businesses

☐ Non-academic research institutions

☐ Nonprofits and community groups

☐ Others

☐ Governmental agencies

☐ Academic institutions

---

#### 11. Evaluation

**Label:** *Evaluation Type*

☐ Effectiveness only

☐ Output only

☐ Both output and effectiveness

☐ No evaluation mentioned

---

#### 12. Evaluation Nature

**Label:** *Nature of Evaluation*

☐ Quantitative—observational

☐ Quantitative—experimental

☐ Qualitative

☐ Not mentioned

☐ Other

---

#### 13. Demographics

**Label:** *Target Population*

☐ Not specified

☐ Children or adolescents

☐ General users

☐ Others

☐ Content creators

☐ Parents

---

#### 14. Instruction

**Label:** *Instruction Type*

☐ Not applicable (by-default design)

☐ User initiation required but not provided

☐ Provide description

☐ Provide link to other pages

☐ Provide image

---

#### 15. Geographic Scope

**Label:** *Geographic Application*

☐ Not specified

☐ Global scale (e.g., applied in 85+ countries)

☐ North America (U.S. & Canada)

☐ EU/UK

☐ Other high-income countries

☐ Low- and middle-income countries

☐ If unsure, specify location below:

---

## **Prompts**

### **Step 1: unit\_of\_analysis\_prompt**

You are a content and qualitative analysis expert.

You will analyze a news announcement from a social media company to identify how many distinct digital well-being designs are described. This task requires careful reasoning, so please follow each step below. Only include designs that satisfy **\*\*BOTH\*\*** of the following inclusion criteria:

**\*\*I1\*\***: The practice must be clearly associated with reducing **\*\*digital harms\*\*** or promoting **\*\*digital well-being\*\***.

This includes, but is not limited to, efforts related to: User safety (e.g., preventing harassment, bullying, or self-harm);

- Privacy and data protection; Exposure to harmful or sensitive content (e.g., violence, disordered eating, hate speech);
- Misinformation and content credibility; Health-related behaviors and risks (e.g., substance use, mental health);
- Age-appropriate design and protections for minors; Encouraging healthier or more mindful usage patterns (e.g., reducing screen time, content fatigue, or compulsive scrolling)

**\*\*I2\*\***: The practice involves a **\*\*clearly defined and implemented product or feature\*\*** that is integrated into the **\*\*platform’s design\*\***. To be included, the change must reflect a concrete design decision made by the platform—that is, something that platform engineers, designers, or product teams have built into the functionality, interface, algorithm, or user flow of the product. This can include: User-facing features (e.g., dashboards, nudges, content warnings, filters, help pop-ups);

- System-level or back-end mechanisms (e.g., algorithmic downranking of harmful content, automated detection of self-harm posts)
- The intervention must go beyond vague or aspirational claims. It should describe a specific, actionable, clearly identifiable and implemented platform-level product or feature—regardless of whether it is new or previously launched.

General statements about values, goals, or unspecified systems should be excluded.

For example, include if the practice is:

- A Help Page is added with links to COVID-19 resources [reasoning: this is a specific platform product (I2) and relate to well-being (I1)]
- The platform rolls out an AI system to detect harmful content targeting on children [reasoning: this is a specific platform feature (I2) and relate to well-being (I1)]

Moreover, Exclude if:

- The platform claims to “prioritize safety” or “aims to reduce harm” without specifying how this is done in the product
- The platform makes general statements like “our systems protect youth” without naming or describing an actual product, feature, or process

If something does not meet both I1 and I2, it must be excluded.

##### Step-by-step instruction:

1. Summarize the overall goals and themes of the announcement. This helps build context. Do not extract any features yet.
2. Identify all statements that could potentially describe a digital well-being design. For each:
  - 2.1) Summarize the statement.
  - 2.2) Evaluate whether it meets I1 and I2.
  - 2.3) If it fails either criterion, explain why and discard it.
3. For the remaining valid candidates, group similar items together if they reflect the same design or concept. Provide justification for each grouping.
4. Count the number of distinct digital well-being designs based on your groupings. Be conservative—if unsure whether two things are distinct, assume they belong to the same category.
5. Provide a final list of distinct digital well-being designs, with the following output format.

##### Output format

If at least one practice qualifies, use this structure:

```
 {"actions": [
 {
 "id": "A1",
 "summary": "Short description of the identified action",
 "reasons": "Explain briefly why action satisfies criteria",
 "uncertainty": "uncertainty score as an integer ranges from 0-5"
 },
 {
 "id": "A2" (if there are more than one practices satisfy both I1 and I2),
 "summary": "Short description of the identified action",
 "reasons": "Explain briefly why action satisfies criteria",
 "uncertainty": "uncertainty score as an integer ranges from 0-5, with 0 means
extremely certain and 5 means extremely uncertain"
 }
],
 "count": number of total practices that match inclusion criteria
}
```

If no changes, then output this format:

```
 {"actions": [],
 "reasons": Brief summarize what is the practice introduced in the announcement (in a short
sentence),
 "uncertainty": "uncertainty score as an integer ranges from 0-5, with 0 means extremely
certain and 5 means extremely uncertain"},
 "count": 0
}
```

Please do not output messages other than the JSON files. ##### The news announcement:

## Step 2: structured\_annotation\_schema\_prompt

You are a content and qualitative analysis expert.

Your task is to analyze a news announcement from a social media company about a digital well-being design using a structured JSON schema. You will be provided with a JSON schema that contains a single **platform-level design practice (PDWP)**, already extracted from the article using prior analysis. This design practice is identified by an id (e.g., "A1") and is summarized in a brief "summary" field. Your goal is to re-analyze the original article and extract structured information strictly based on the article content. Do not generate annotations based on inference.

##### Step-by-step instruction:

1. Extract a block of sentence(s) from the news announcement that:

- Clearly describe the design change itself, and
- Provide important context that helps explain why, for whom, or how it was introduced

This block should include:

- Sentences that directly describe the design
- Nearby sentences that clarify implementation, outcomes, or target audience
- Summary or framing sentences from the earlier introduction or elsewhere in the article, \ if they help explain the problem, harm, target, or motivation for this design

In other words: extract the **full contextual span** of the PDWP, even if part of the context appears outside of the immediate paragraph (such as in the article's opening summary).

2. Populate the "annotations" field selectively

After extracting the full contextual block for the given PDWP, examine it carefully to identify whether it contains any explicit supporting details relevant to the annotation fields. Use the entire contextual span, including any background or framing sentences you included in Step 1, as your source of evidence. Do not guess or infer. Only include an annotation if the article clearly and explicitly provides it. For each field in the 'annotations', ask yourself:

"Does the news announcement explicitly mention this for this specific design change?"

If yes → include the field, using a concise and faithful phrase or short sentence.

If no → omit the field entirely.

##### Annotation Fields You Should Check (Only Include When the Field is Mentioned for the Specific PDWP):

Field	Meaning
-----	-----
----	----
`explicit_harm`	If the news announcement described the PDWP addresses digital harm or user well-being, summarize it
`entities`	If the news announcement named an external organizations, companies, or individuals in the text around the specific PDWP, extract all names of the mentioned entities.

Please do not miss external organization names, even if it is already extracted in other fields such as "evaluation", "partnership".

| `motivation` | If the announcement says why the PDWP is introduced or the background motivation of introducing the PDWP (e.g., "to help prevent harm", "to address abuse", "to improve user experience", "follow community feedback", "follow specific events"), summarize the motivations |

| `motivation\_evidence` | If the news announcement mentions or cites studies, user feedback, data, or other materials in the background of explaining why the PDWP is being introduced or what motivates its provision, summarize that evidence. This includes both quantitative evidence (e.g., internal studies, metrics, testing) and qualitative evidence (e.g., user interviews, pilot feedback, or qualitative assessments). Pay special attention to any quantitative numbers that indicate measurable outcomes or impact, and include these numbers in the summary when available. |

| `partnership` | If the news announcement described specific collaborators or targets (e.g., universities, nonprofits) involved in the PDWP, extract all names of the mentioned collaborators |

| `geographic\_scope` | If the news announcement stated that the PDWP applies to a specific country, region, or market, list all names of the regions that the PDWP will be applied to |

| `demographics` | If the news announcement stated that the PDWP targets a user group (e.g., teens, parents, creators, or all users), list all names of the groups that the PDWP will be applied to |

| `evaluation` | If the news announcement mentions any impact measures, output metrics, or evaluation efforts related to the PDWP's output, outcome, or effectiveness, summarize them. This includes both quantitative evaluations (e.g., internal studies, metrics, testing) and qualitative efforts (e.g., user interviews, pilot feedback, or qualitative assessments). Pay special attention to any quantitative numbers that indicate measurable outcomes or impact, and include these numbers in the summary when available. |

| `instruction` | If the news announcement described how to use, access, or interact with the PDWP, summarize the instruction |

You should include only the fields for which there is explicit, article-based evidence. If a field is not mentioned, omit it entirely.

#### #### Output Format

```
{{
 "A1": {{
 "extractive": "Verbatim sentence(s) from the news announcement that you extracted for the
specific PDWP",
 "annotations": {{
 "explicit_harm": "string (only if the news announcement mentions it)",
 "entities": "string (only if the news announcement mentions it)",
 "motivation": "string (only if the news announcement mentions it)",
 "motivation_evidence": "string (only if the news announcement mentions it)",
 "partnership": "string (only if the news announcement mentions it)",
```

```

 "geographic_scope": "string (only if the news announcement mentions it)",
 "demographics": "string (only if the news announcement mentions it)",
 "evaluation": "string (only if the news announcement mentions it)",
 "instruction": "string (only if the news announcement mentions it)"
 }}
}}

```

### ***Step 3: Prompts for inferential classification***

```

anno_codebook = {
 'explicit_harm':[
 'behavioral harm (e.g., long hours of use, sleep disturbance)',
 'harmful content',
 'harmful contact (e.g., unwanted direct message)',
 'privacy-related concerns',
 'transparency and accountability (e.g., rules of prohibited content)',
 'external harm (e.g., COVID, suicidality)'
],

 'demographics':[
 'child or adolescent',
 'parents',
 'content creators',
 'general users',
 'others (please specify in the output)'
],

 'motivation': [
 'research',
 'civil society',
 'legislation',
 'community pressure',
 'advocacy groups',
 'events (e.g., World Eating Disorder Action Day)',
 'company values',
 'others (please specify in the output)'
],

 'motivation_evidence': [
 'quantitative-observational',
 'quantitative-experimental',
 'qualitative',
 'other',
]
}

```

```

 'not mentioned'
],

 'motivation_nature': [
 'quantitative-observational',
 'quantitative-experimental',
 'qualitative',
 'other',
 'not mentioned'
],

 'geographic_scope': [
 'List full country, state, or region names, separated by commas'
],

 'instruction': [
 'provide url',
 'provide image',
 'provide description',
 'do not provide'
],

 'evaluation': [
 'Output only (defined as what the change produces, e.g., \
a tool used by 16 million users or some internal team made 10 different tools.)',
 'Effectiveness only (defined as the effect of the PDWP on the desired result, e.g., \
improve proportion of users raising mental health questions by 5%, or reduce 75%
misinformation.)',
 'Both output and effectiveness'
],

 'evaluation_nature': [
 'quantitative-observational',
 'quantitative-experimental',
 'qualitative',
 'other',
 'not mentioned'
],

 'partnership': [
 'academic institution',
 'non-academic research institution',
 'nonprofits and other community groups',
 'governmental agencies',

```

```

 'businesses',
 'others (please specify in the output)'
],

 'entities': [
 'academic institution',
 'non-academic research institution',
 'nonprofits and other community groups',
 'governmental agencies',
 'businesses',
 'others (please specify in the output)'
],

 'automaticity': ['by default',
 'user-initiated'
],

 'design_element':[
 "Infinite Scroll",
 "Auto Play",
 "Engagement Gamification",
 "Notifications (Platform and User Content)",
 "Requiring Account Creation",
 "Ephemeral Content",
 "Barriers to Deletion",
 "Attention Optimization",
 "Personalization",
 "Lack of Rate Limits",
 "Lack of Negative Feedback (Barriers to Reporting)",
 "Badges and Verification",
 "Discriminatory Algorithms",
 "Profile / Picture Effects",
 "Generative AI Platform Tools",
 "Generative AI User Tools",
 "Lack of Friend/Connection/Messaging Rate Limits",
 "Default Profile Privacy",
 "Default Post/Comment Privacy",
 "User Discoverability (Search or Friending)",
 "Age Verification",
 "Geolocation",
 "Disproportionate Data Collection by Platform"
]
}

```

```

question_codebook = {
'explicit_harm': "According to the **explicit_harm summary**, which harm category (or
categories) in the **annotation codebook** best describe the type of harm the PDWP is
intended to address?",

'demographics': "According to the **demographics summary**, which demographics category
(or categories) in the **annotation codebook** best describe the type of target users of the
PDWP?",

'motivation': "According to the **motivation summary**, which type of motivation(s) provided in
the **annotation codebook** best explain why the platform introduced the PDWP?",

'motivation_evidence': "According to the **motivation_evidence summary**, what type of
evidence in the **annotation codebook** was provided about what motivate the platform to
initiate the PDWP (if any)?",

'motivation_nature': "According to the **motivation_nature summary**, what type of evidence in
the **annotation codebook** was provided about what motivate the platform to initiate the
PDWP (if any)?",

'geographic_scope': "According to the **geographic_scope summary**, what geographic regions
(country, state, or area) does the PDWP apply to? Please list full names.",

'instruction': "According to the **instruction summary**, what type of instruction or guidance in
the **annotation codebook** is provided by the PDWP (if any)?",

'evaluation': "According to the **evaluation summary**, what type of evaluation results provided
in the **annotation codebook** was mentioned in the PDWP?",

'evaluation_nature': "According to the **evaluation_nature summary**, what type of evaluation
evidence in the **annotation codebook** was provided about the PDWP (if any)?",

'partnership': "According to the **partnership summary**, what types of partners provided in the
annotation codebook were mentioned in carrying out the PDWP?",

'entities': "Which entity category (or categories) from the **annotation codebook** best describe
the entity (or entities) mentioned in the **entities summary**?",

"automaticity": "Does it indicate that the change is: 1) applied by default and does not require
user action, or 2) Requires users to apply, initiate, or access it?",

'design_element': "Which design element(s) from the list are referenced or implicated in the
summary?"

```

```
}
```

```
evidence and evaluation nature
```

```
def generate_minimal_prompt(news_extracive, news_anno_summary, annotation_field,
annotation_field_code, question):
 codebook_text = "\n".join([f"- {item}" for item in annotation_field_code])
```

```
 prompt = textwrap.dedent(f""" You are a content and qualitative analysis expert. Your task is
to assign classification tags to a **platform-level design well-being practices (PDWP)** using a
structured annotation codebook. The PDWP represents a platform-level design feature intended
to promote user well-being or reduce user harms on a social media platform.
```

```
Read the following **extractive** and the **summary** about the PDWP's **{annotation_field}**
carefully.
```

```
Extractive: {news_extracive}
```

```
{annotation_field.capitalize()} Summary: {news_anno_summary}
```

```
Given the following **annotation codebook** of `{annotation_field}`:
{codebook_text}
```

```
And given the extractive and annotation summary above, choose one or more categories from
the annotation codebook of `{annotation_field}` that best answer the following **question**:
{question}
```

```
Do not infer, generalize, or assume information beyond what the annotations support. Assign
multiple tags within a field only when clearly justified by the annotations. When choosing tags for
the annotation codebook, only include the main phrase, no need to include the examples in the
parentheses in the final JSON output.
```

```
Output format:
```

```
``json
```

```
{{
```

```
 "{annotation_field}": ["tag1", "tag2" (only if there are more than one categories)]
```

```
}}
```

```
Please do not output messages other than the JSON files.
```

```
""")
```

```
 return prompt
```

```
def generate_minimal_prompt_extrac(news_extracive, annotation_field, annotation_field_code,
question):
 codebook_text = "\n".join([f"- {item}" for item in annotation_field_code])
```

prompt = textwrap.dedent(f""" You are a content and qualitative analysis expert. Your task is to assign classification tags to a **platform-level design well-being practices (PDWP)** using a structured annotation codebook. The PDWP represents a platform-level design feature intended to promote user well-being or reduce user harms on a social media platform.

Read the following **summary** about the PDWP's **{annotation\_field}** carefully.  
**summary**: {news\_extractive}

Given the following **annotation codebook** of **{annotation\_field}**:  
{codebook\_text}

And given the **summary** above, choose one or more categories from the annotation codebook of **{annotation\_field}** that best answer the following **question**:  
**{question}**

Do not infer, generalize, or assume information beyond what the annotations support. Assign multiple tags within a field only when clearly justified by the annotations. When choosing tags for the annotation codebook, only include the main phrase, no need to include the examples in the parentheses in the final JSON output.

### Output format:

```
```json
{{
  "{annotation_field}": ["tag1", "tag2" (only if there are more than one categories)]
}}
```

Please do not output messages other than the JSON files.

```
""")
return prompt
```

Step 4: Prompt for the thematic analysis

4.1: summarize design features

You are an expert in content thematic analysis. Your task is to extract a concise **Action code** — a short phrase that clearly describes what the social media platform is doing or implementing — based on a description of **platform-level design practices (PDP)** on a social media platform.

The PDP represents a **concrete platform-level design feature or initiative**. Focus only on the **specific action or intervention taken by the platform**, not general context, opinion, or background. Your output should be a neutral, descriptive phrase that summarizes the core platform action.

Data format

Corpus: [original description you will receive from us]

Action code: [short summary of what the platform is doing that you will output to us]

Please see the examples below:

****Example 1****

Corpus: A couple of weeks ago we talked about how, creators need to start disclosing in YouTube Studio desktop if the content we're uploading has been altered or synthetically generated in a meaningful way. Now, in addition to the YouTube blog post and help center, I had the chance to sit down and chat with our VP of product management, Jennifer Flannery O'Connor, to talk about not just how the disclosure and labeling works for creators and viewers, but why it's needed, what the guiding principles are behind it, and how it might evolve from here.:

Action code: AI-generated content disclosure for creators

****Example 2****

Corpus: We've also heard from creators about the importance of broadening accessibility for people who are deaf or hard of hearing. Thanks to long term investments in machine learning, we now provide automatic captions in more languages. We also offer machine translated captions for mobile that enable viewers to translate their captions to 16 languages. YouTube has captioned over six billion videos with more than one billion users watching videos with captions enabled every day.

Action code: Expanded automatic and translated captions for accessibility.

****Example 3****

Corpus: We have also added new measures to limit Snapchatters under 18 from showing up in search results or as a friend suggestion to someone else unless they have friends in common. This builds on protections we've long had in place that require teens to be friends with another Snapchat user in order to communicate directly.

Action code: Limit teen visibility in search and friend suggestions

****Example 4****

Corpus: Today we are also releasing new in-app content that helps explain these risks and shares important resources for Snapchatters, such as hotlines to contact for help. This content will be featured on our Stories platform and surfaced to Snapchatters by relevant Search terms or keywords. Several of our new product safeguards were informed by feedback from The National Center on Sexual Exploitation (NCOSE). Our new in-app educational resources were developed with The National Center for Missing and Exploited Children (NCMEC).

Action code: In-app educational resources

****Output format****

Output only the ****Action code****. Do not include any explanation or interpretation.

Please work on:

Corpus: {extractive}

Action code:

4.2: summarize harms

You are an expert in content thematic analysis. Your task is to extract a concise ****Well-being topic code**** — a short phrase that clearly identifies the specific well-being-related issue, harm, or target group mentioned in a description of platform-level digital well-being design practices (PDWP) from a social media platform. If the content is not in English, please also translate your response to English.

A PDWP refers to a platform-level design feature or intervention aimed at promoting user well-being or reducing harm.

Only include a well-being topic if it is explicitly mentioned in the text. Do not infer, generalize, or interpret beyond the given content.

Your output should be a neutral, descriptive phrase summarizing the well-being issue in English. If no well-being-related topic is clearly mentioned, write: "None".

****Data format****

Corpus: [original description you will receive from us]

Well-being topic code: [short phrase or "None"]

Please see the examples below:

****Example 1****

Corpus: A couple of weeks ago we talked about how, creators need to start disclosing in YouTube Studio desktop if the content we're uploading has been altered or synthetically generated in a meaningful way. Now, in addition to the YouTube blog post and help center, I had the chance to sit down and chat with our VP of product management, Jennifer Flannery O'Connor, to talk about not just how the disclosure and labeling works for creators and viewers, but why it's needed, what the guiding principles are behind it, and how it might evolve from here.:

Well-being topic code: None

****Example 2****

Corpus: We've also heard from creators about the importance of broadening accessibility for people who are deaf or hard of hearing. Thanks to long term investments in machine learning, we now provide automatic captions in more languages. We also offer machine translated captions for mobile that enable viewers to translate their captions to 16 languages. YouTube has captioned over six billion videos with more than one billion users watching videos with captions enabled every day.

Well-being topic code: deaf people

****Example 3****

Corpus: We have also added new measures to limit Snapchatters under 18 from showing up in search results or as a friend suggestion to someone else unless they have friends in common. This builds on protections we've long had in place that require teens to be friends with another Snapchatter in order to communicate directly.

Well-being topic code: Adolescent privacy

****Example 4****

Corpus: Today we are also releasing new in-app content that helps explain these risks and shares important resources for Snapchatters, such as hotlines to contact for help. This content will be featured on our Stories platform and surfaced to Snapchatters by relevant Search terms or keywords. Several of our new product safeguards were informed by feedback from The National Center on Sexual Exploitation (NCOSE). Our new in-app educational resources were developed with The National Center for Missing and Exploited Children (NCMEC).

Well-being topic code: Sexual Exploitation and exploited children

****Output format****

Output only the ****Well-being topic code**** [None or the generated code]. Do not include any explanation or interpretation. Please translate the response to English\

Do not include explanations or reasoning. Do not generate Well-being topic code unless a well-being issue is explicitly stated.

Please work on the following:

Corpus: {extractive}

Well-being topic code:

4.3: merge themes

You are an expert in iterative content thematic analysis. I will provide a list of short, descriptive ****initial codes**** that summarize ****platform-level digital design practices (PDP)**** from a large dataset. Each code reflects the action described in a corpus item.

Your task is to group similar or related codes into broader well-defined themes. Each theme should represent a meaningful pattern or design category.

Please generate between 15-30 themes.

Avoid overly narrow, redundant, or overlapping themes by merging where appropriate.

If any code does not clearly belong to a theme, include it in an "Uncategorized" group.

**** Input data format **:**

code 1: [example code text]

code 2: [example code text]

...

code 50: [example code text]

**** output format ****

Return the output as a JSON dictionary with the following structure:

```
{
  "T1": {
    "theme_label": "short label summarizing the theme",
    "theme_description": "brief explanation of what the theme captures",
    "theme_example": [
      "example initial code 1",
      "example initial code 2",
      "example initial code 3",
      "example initial code 4", # if there is a fourth one
    ]
  },
  "T2": {
    "theme_label": "...",
    "theme_description": "...",
    "theme_example": [...],
  },
  ...,
  "Uncategorized": {
    "example": [
      "initial code not categorized",
      "another uncategorized code"
    ]
  },
}
```

Now it's your turn. Below is the input:

(Do not include any explanation or text outside of the JSON output.)

{text}

4.4: text classification

You are an expert in iterative content thematic analysis.

Your task is to assign the most appropriate ****theme label(s)**** to the given ****corpus item****, based on a finalized ****theme codebook**** developed through prior rounds of analysis.

The ****theme codebook**** defines the thematic categories within a taxonomy of platform design features.

Each ****corpus item**** represents a specific platform-level design practice implemented by a social media platform.

Some ****corpus item**** may clearly reflect one or more design features captured in the ****theme codebook****, while others may not fit any theme. Note that some ****corpus item**** may not be in English, please return the theme label(s) in English.

Your goal is to carefully read each **corpus item** and assign only the most relevant theme label(s) from the theme codebook.

Classification guide

Use only the theme labels provided in the codebook.

If the corpus does not match any existing theme, label it as "None".

Each corpus may be assigned zero, one, or multiple theme labels.

Output format

Please return your result as plain text, listing only the assigned theme label(s).

If one or more themes apply, return the exact theme labels separated by commas (for example, "User Content Control,Misinformation Detection and Reduction")

If no theme applies, return the word "Uncategorized".

Do not include quotation marks, brackets, or any explanation, but just the clean text.

Now it's your turn. Below is the **theme codebook** and **corpus item**:

Theme Codebook

Each theme includes a label, a brief description, and several examples. You must use these themes exactly as provided for classification.

{theme_codebook}

Corpus Item

{text}

Chapter 5: Conclusion

5.1 Key Findings and Implications

This dissertation advances research on social media use and well-being from three perspectives: by developing scalable objective behavioral measurement through user-donated screen time screenshots and extending the application of data donation methods using TikTok trace data; by shifting the focus from specific types of use to a cross-cutting psychological process that conceptualizes social media engagement as momentary goal pursuit; and by examining platform-level digital well-being practices to identify gaps in transparency and inform policy discussions.

Chapter 2 demonstrates a scalable approach to capturing objective, app-specific social media behaviors through user-donated screenshots and highlights the importance of examining both between- and within-person variation in the relationship between social media use and adolescent well-being. Overall, the results showed predominantly small or non-significant associations between app-specific usage metrics and both hedonic and eudaimonic well-being outcomes. The findings suggest that focusing solely on app-level usage may offer limited insight into the mechanisms linking social media use and well-being, underscoring the importance of considering broader psychological and contextual factors.

Chapter 3 shifts the focus from specific types of social media use to the underlying psychological process of goal pursuit, demonstrating that higher levels of momentary goal awareness and goal strength are associated with more positive and meaningful experiences. The findings highlight the potential of incorporating goal awareness into user-centered digital literacy interventions. At the same time, we raise important theoretical questions about how habitual use can be meaningfully integrated into goal-pursuit frameworks within digital environments. The limited associations between goal awareness, goal strength, and objective TikTok usage patterns further underscore the complexity of real-world digital behavior, suggesting the need for more

refined study designs and analytical approaches to better capture high-dimensional engagement data.

Chapter 4 broadens the lens to the platform level by documenting the expansion of platform-reported digital well-being practices through analysis of company announcements. Although platforms increasingly signal a commitment to digital well-being, these commitments largely emphasize content removal, information provision, and optional user controls. The announcements provide limited evidence regarding the effectiveness of these practices and rarely report structural changes to default settings or algorithmic systems. These findings suggest that reliance on newsroom disclosures is insufficient as a robust mechanism, underscoring the need for more rigorous, evidence-based, and transparent frameworks for monitoring platform digital well-being practices.

Overall, the empirical studies in this dissertation are subject to sampling constraints that limit the generalizability of their findings. Chapter 2 recruited Swiss adolescents through high schools, and Chapter 3 recruited U.S. adults via Prolific with screening criteria targeting frequent TikTok users. Neither sample is representative of a broader national or international population. In both studies, participants who consented to share behavioral data may differ systematically from those who declined, for example in their comfort with data sharing, level of platform engagement, or general orientation toward social media research, introducing an additional layer of selection bias that constrains inference from the objective behavioral analyses in particular. I want to note that these sampling constraints are not incidental but reflect a broader structural challenge in digital well-being research: the populations most accessible to researchers — young, educated, Western, and digitally active — are precisely those for whom existing evidence is already most abundant.

5.2 Future Research

Building on the empirical chapters, I outline several directions for future research. First, following Chapter 2, I am developing a framework to examine the variability of individuals' social media use using objective data, capturing patterns such as fragmentation, routine, complexity, and uncertainty beyond time spent. Second, extending Chapter 3, I integrate objective TikTok logs with multimodal content data from short-form TikTok videos and interaction behaviors to identify content consumption patterns associated with goal awareness, goal strength, and well-being outcomes. I also develop a taxonomy for analyzing short-form social media videos, outlining multiple dimensions of analysis, including topic, sentiment, stylistic and structural features, and mid-level expressive qualities such as tone, pacing, and overall "vibe," to enable more systematic and scalable content analysis. Third, building on Chapter 4, I expand the analysis of digital well-being practices to AI companion chatbots, combining platform disclosures with user discussions on forums to better understand user experiences in emerging AI-mediated environments.

An ongoing challenge in the study of digital well-being is the tension between relatively modest sample sizes and increasingly high-dimensional behavioral data. Related to this is the broader question of what constitutes a meaningful effect size in this domain, a question the field has yet to resolve. Conventional benchmarks such as Cohen's $d = 0.2$ [316] were derived from between-person, cross-sectional designs and are not directly applicable to within-person EMA analyses, where variance partitioning at the person level inherently produces smaller standardized coefficients. Rigorous longitudinal work in social media use and well-being has found that within-person effects of social media use on life satisfaction are trivial in size, suggesting that small effects are characteristic of the broader literature rather than indicative of methodological inadequacy [101]. At the same time, small within-person effects should not be dismissed, when consistent and cumulative across the millions of users who engage with social media daily, even modest associations may carry substantial public health implications at scale, underscoring the importance of replication, causal identification, and app-specific precision in

future work. While practical and ethical constraints often limit large-scale data collection, future research would benefit from substantially larger samples. Emerging large-scale data-donation initiatives [71] and regulatory developments (e.g., Digital Services Act [24]) may help facilitate broader data access, enabling more scalable and reproducible investigations.

Beyond social media, the empirical findings and conceptual insights from this dissertation offer relevant implications for emerging AI systems, such as companion chatbots and AI tools. Like social media, AI technologies are rapidly becoming embedded in daily life, characterized by personalization, persistent interaction, and the potential to shape users' psychological and behavioral processes [317–319]. Current research on AI engagement often relies on self-reported usage and attitude, and focuses on identifying benefits or harms. These approaches are similar to those used in early social media studies. The lessons learned from social media research, particularly regarding objective behavioral measurement, user agency, and goal pursuit, and platform governance on user safety, therefore provide a valuable foundation for studying AI systems.

Looking ahead, extending the study of digital well-being to more diverse and underrepresented populations, such as those in the Global South, represents both a longer-term goal of this research and a broader priority for the field [320]. Much of the existing evidence on social media and digital well-being relies on online survey samples or students from high-income countries, where stronger data infrastructures and research resources facilitate data collection. Objective behavioral data and EMA studies on digital well-being remain far less common in middle- and low-income contexts. As a result, current knowledge on digital well-being is disproportionately derived from high-income contexts—a form of data absenteeism [321], the systematic absence of data on groups experiencing social vulnerability—limiting its generalizability, which also reflects broader structural inequalities in global research capacity. Challenges related to social media and AI use may differ across socio-economic, cultural, and regulatory environments. Expanding this research to the Global South is therefore essential not

only for representativeness but also for identifying context-specific risks, opportunities, and governance needs. Addressing this gap would contribute to a more equitable and globally informed framework for digital well-being.

Reference

1. Schrape JF. Social media, mass media and the “public sphere”: differentiation, complementarity and co-existence. Vols. 2016–01. Stuttgart: Universität Stuttgart, Fak. 10 Wirtschafts- und Sozialwissenschaften, Institut für Sozialwissenschaften Abt. VI Organisations- und Innovationssoziologie; 2016. 24 p. (Research contributions to organizational sociology and innovation studies / Stuttgarter Beiträge zur Organisations- und Innovationssoziologie : SOI discussion paper).
2. Humphreys L. Mobile social media: The challenges and opportunities continue. *Mobile Media & Communication*. 2023 Jan 1;11(1):74–9. doi:10.1177/20501579221132209
3. Kristian Lum, Tomo Lazovich. The Myth of The Algorithm: A system-level view of algorithmic amplification [Internet]. 2023 [cited 2024 Apr 11]. Available from: <http://knightcolumbia.org/content/the-myth-of-the-algorithm-a-system-level-view-of-algorithmic-amplification>
4. Bhargava VR, Velasquez M. Ethics of the Attention Economy: The Problem of Social Media Addiction. *Business Ethics Quarterly*. 2021 Jul;31(3):321–59. doi:10.1017/beq.2020.32
5. Mujica AL, Crowell CR, Villano MA, Uddin KM. ADDICTION BY DESIGN: Some Dimensions and Challenges of Excessive Social Media Use. *Medical Research Archives*. 2022 Feb 24;10(2). doi:10.18103/mra.v10i2.2677
6. Arnold M, Goldschmitt M, Rigotti T. Dealing with information overload: a comprehensive review. *Front Psychol*. 2023 Jun 21;14. doi:10.3389/fpsyg.2023.1122200
7. Du J, Kerkhof P, van Koningsbruggen GM. Predictors of Social Media Self-Control Failure: Immediate Gratifications, Habitual Checking, Ubiquity, and Notifications. *Cyberpsychology, Behavior, and Social Networking*. 2019 Jul;22(7):477–85. doi:10.1089/cyber.2018.0730
8. Muhammed T S, Mathew SK. The disaster of misinformation: a review of research in social media. *Int J Data Sci Anal*. 2022 May 1;13(4):271–85. doi:10.1007/s41060-022-00311-6
9. WHO. Over a billion people living with mental health conditions – services require urgent scale-up [Internet]. 2025 [cited 2026 Feb 12]. Available from: <https://www.who.int/news/item/02-09-2025-over-a-billion-people-living-with-mental-health-conditions-services-require-urgent-scale-up>
10. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry*. 2022 Feb 1;9(2):137–50. doi:10.1016/S2215-0366(21)00395-3

11. WHO. Teens, screens and mental health [Internet]. [cited 2026 Feb 12]. Available from: <https://www.who.int/europe/news/item/25-09-2024-teens--screens-and-mental-health>
12. Tian J, Yan N, Hu X, Tian S, Wang Y, Mackay LE, et al. Global burden of mental disorders among adolescents and young adults, 1990–2021: a systematic analysis of the Global Burden of Diseases Study 2021. *Gen Psychiatr*. 2025 Dec 23;38(6):e102278. doi:10.1136/gpsych-2025-102278 PubMed PMID: 41451038; PubMed Central PMCID: PMC12730781.
13. Leeb RT, Danielson ML, Claussen AH, Robinson LR, Lebrun-Harris LA, Ghandour R, et al. Trends in Mental, Behavioral, and Developmental Disorders Among Children and Adolescents in the US, 2016–2021. *Prev Chronic Dis*. 2024 Dec 12;21:E96. doi:10.5888/pcd21.240142 PubMed PMID: 39666872; PubMed Central PMCID: PMC11640887.
14. Motyl M. Measuring Changing Social Media Experiences Across 2 Years [Substack newsletter]. *Designing Tomorrow* [Internet]. 2025 Apr 24 [cited 2026 Feb 12]. Available from: <https://psychoftech.substack.com/p/social-media-experiences-across-nearly>
15. Weinstein E. The social media see-saw: Positive and negative influences on adolescents' affective well-being. *New Media & Society*. 2018 Oct 1;20(10):3597–623. doi:10.1177/1461444818755634
16. Davis K, Charmaraman L. Introduction to the Section on Youth Digital Wellbeing. In: Christakis DA, Hale L, editors. *Handbook of Children and Screens: Digital Media, Development, and Well-Being from Birth Through Adolescence* [Internet]. Cham: Springer Nature Switzerland; 2025 [cited 2026 Feb 12]. p. 211–4. Available from: https://doi.org/10.1007/978-3-031-69362-5_29 doi:10.1007/978-3-031-69362-5_29
17. Advisory of U.S. Surgeon General. Social Media and Youth Mental Health — Current Priorities of the U.S. Surgeon General [Internet]. 2023 [cited 2024 Apr 8]. Available from: <https://www.hhs.gov/surgeongeneral/priorities/youth-mental-health/social-media/index.html>
18. Kross E, Verduyn P, Sheppes G, Costello CK, Jonides J, Ybarra O. Social Media and Well-Being: Pitfalls, Progress, and Next Steps. *Trends Cogn Sci*. 2021 Jan;25(1):55–66. doi:10.1016/j.tics.2020.10.005 PubMed PMID: 33187873.
19. Orben A. Teenagers, screens and social media: a narrative review of reviews and key studies. *Soc Psychiatry Psychiatr Epidemiol*. 2020 Apr 1;55(4):407–14. doi:10.1007/s00127-019-01825-4
20. Valkenburg PM. Social media use and well-being: What we know and what we need to know. *Current Opinion in Psychology*. 2022 Jun 1;45:101294. doi:10.1016/j.copsyc.2021.12.006

21. Al-Mansoori RS, Al-Thani D, Ali R. Designing for Digital Wellbeing: From Theory to Practice a Scoping Review. *Human Behavior and Emerging Technologies*. 2023;2023(1):9924029. doi:10.1155/2023/9924029
22. Alex Moehring, Alissa Cooper, et al. Better Feeds: Algorithms That Put People First. Knight-Georgetown Institute [Internet]. 2025 Mar 4 [cited 2026 Feb 12]. Available from: <https://kgi.georgetown.edu/research-and-commentary/better-feeds/>
23. General Data Protection Regulation (GDPR) [Internet]. [cited 2025 Oct 23]. General Data Protection Regulation (GDPR) – Legal Text. Available from: <https://gdpr-info.eu/>
24. DSA order. The Digital Services Act package | Shaping Europe’s digital future [Internet]. 2024 [cited 2024 May 20]. Available from: <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>
25. Nicol Turner Lee, Josie Stewart, Carolina Oxenstierna. How will bans on social media affect children? Brookings [Internet]. 2025 Dec 9 [cited 2026 Feb 12]. Available from: <https://www.brookings.edu/articles/how-will-bans-on-social-media-affect-children/>
26. Orben A. Digital diet: A 21st century approach to understanding digital technologies and development. *Infant and Child Development*. 2022;31(1):e2228. doi:10.1002/icd.2228
27. High AC, Ruppel EK, McEwan B, Caughlin JP. Computer-Mediated Communication and Well-Being in the Age of Social Media: A Systematic Review. *Journal of Social and Personal Relationships*. 2023 Feb 1;40(2):420–58. doi:10.1177/02654075221106449
28. Meier A, Reinecke L. Computer-Mediated Communication, Social Media, and Mental Health: A Conceptual and Empirical Meta-Review. *Communication Research*. 2021 Dec 1;48(8):1182–209. doi:10.1177/0093650220958224
29. Appel M, Marker C, Gnambs T. Are Social Media Ruining Our Lives? A Review of Meta-Analytic Evidence. *Review of General Psychology*. 2020 Mar 1;24(1):60–74. doi:10.1177/1089268019880891
30. Marciano L, Lin J, Sato T, Saboor S, Viswanath K. Does social media use make us happy? A meta-analysis on social media and positive well-being outcomes. *SSM - Mental Health*. 2024 Dec 1;6:100331. doi:10.1016/j.ssmmh.2024.100331
31. Verduyn P, Ybarra O, Résibois M, Jonides J, Kross E. Do Social Network Sites Enhance or Undermine Subjective Well-Being? A Critical Review. *Social Issues and Policy Review*. 2017;11(1):274–302. doi:10.1111/sipr.12033
32. Wang JL, Gaskin J, Rost DH, Gentile DA. The Reciprocal Relationship Between Passive Social Networking Site (SNS) Usage and Users’ Subjective Well-Being. *Social Science Computer Review*. 2018 Oct 1;36(5):511–22. doi:10.1177/0894439317721981
33. Escobar-Viera CG, Shensa A, Bowman ND, Sidani JE, Knight J, James AE, et al. Passive and Active Social Media Use and Depressive Symptoms Among United States Adults.

- Cyberpsychology, Behavior, and Social Networking. 2018 Jul;21(7):437–43.
doi:10.1089/cyber.2017.0668
34. Meier A, Krause HV. Does Passive Social Media Use Harm Well-Being? *Journal of Media Psychology* [Internet]. 2022 Oct 20 [cited 2026 Jan 4]. Located at: world. Available from: <https://econtent.hogrefe.com/doi/10.1027/1864-1105/a000358>
 35. Godard R, Holtzman S. Are active and passive social media use related to mental health, wellbeing, and social support outcomes? A meta-analysis of 141 studies. *J Comput Mediat Commun*. 2024 Jan 1;29(1):zmad055. doi:10.1093/jcmc/zmad055
 36. Sayag M, Most C, Katzir M, Luria R, Kross E, Sheppes G. Posting on Social-Media Undermines Academic Performance and Well-Being [Internet]. *Research Square*; 2024 [cited 2026 Jan 4]. Available from: <https://www.researchsquare.com/article/rs-3805219/v1> doi:10.21203/rs.3.rs-3805219/v1
 37. Karsay K, Matthes J, Schmuck D, Ecklebe S. Messaging, Posting, and Browsing: A Mobile Experience Sampling Study Investigating Youth’s Social Media Use, Affective Well-Being, and Loneliness. *Social Science Computer Review*. 2023 Aug 1;41(4):1493–513. doi:10.1177/08944393211058308
 38. Bailey ER, Matz SC, Youyou W, Iyengar SS. Authentic self-expression on social media is associated with greater subjective well-being. *Nat Commun*. 2020 Oct 6;11(1):4889. doi:10.1038/s41467-020-18539-w
 39. Burnell K, George MJ, Underwood MK. Browsing Different Instagram Profiles and Associations With Psychological Well-Being. *Front Hum Dyn*. 2020 Nov 19;2. doi:10.3389/fhumd.2020.585518
 40. Satıcı SA, Gocet Tekin E, Deniz ME, Satıcı B. Doomscrolling Scale: its Association with Personality Traits, Psychological Distress, Social Media Use, and Wellbeing. *Applied Research Quality Life*. 2023 Apr 1;18(2):833–47. doi:10.1007/s11482-022-10110-7
 41. Terzimehic N, Bemmam F, Halsner M, Mayer S. A Mixed-Method Exploration into the Mobile Phone Rabbit Hole. *Proc ACM Hum-Comput Interact*. 2023 Sep 11;7(MHCI):1–29. doi:10.1145/3604241
 42. Buchanan K, Aknin LB, Lotun S, Sandstrom GM. Brief exposure to social media during the COVID-19 pandemic: Doom-scrolling has negative emotional consequences, but kindness-scrolling does not. *PLOS ONE*. 2021 Oct 13;16(10):e0257728. doi:10.1371/journal.pone.0257728
 43. Cauberghe V, Van Wesenbeeck I, De Jans S, Hudders L, Ponnet K. How Adolescents Use Social Media to Cope with Feelings of Loneliness and Anxiety During COVID-19 Lockdown. *Cyberpsychology, Behavior, and Social Networking*. 2021 Apr;24(4):250–7. doi:10.1089/cyber.2020.0478

44. Verma G, Bhardwaj A, Aledavood T, De Choudhury M, Kumar S. Examining the impact of sharing COVID-19 misinformation online on mental health. *Sci Rep.* 2022 May 16;12(1):8045. doi:10.1038/s41598-022-11488-y
45. Saha K, Chandrasekharan E, De Choudhury M. Prevalence and Psychological Effects of Hateful Speech in Online College Communities. In: *Proceedings of the 10th ACM Conference on Web Science* [Internet]. New York, NY, USA: Association for Computing Machinery; 2019 [cited 2026 Jan 5]. p. 255–64. (WebSci '19). Available from: <https://doi.org/10.1145/3292522.3326032> doi:10.1145/3292522.3326032
46. Giumetti GW, Kowalski RM. Cyberbullying via social media and well-being. *Current Opinion in Psychology.* 2022 Jun 1;45:101314. doi:10.1016/j.copsyc.2022.101314
47. Dane A, Bhatia K. The social media diet: A scoping review to investigate the association between social media, body image and eating disorders amongst young people. *PLOS Global Public Health.* 2023 Mar 22;3(3):e0001091. doi:10.1371/journal.pgph.0001091
48. Farrell T, Fernandez M, Novotny J, Alani H. Exploring Misogyny across the Manosphere in Reddit. In: *Proceedings of the 10th ACM Conference on Web Science* [Internet]. New York, NY, USA: Association for Computing Machinery; 2019 [cited 2026 Jan 5]. p. 87–96. (WebSci '19). Available from: <https://doi.org/10.1145/3292522.3326045> doi:10.1145/3292522.3326045
49. Naslund JA, Aschbrenner KA, McHugo GJ, Unützer J, Marsch LA, Bartels SJ. Exploring opportunities to support mental health care using social media: A survey of social media users with mental illness. *Early Intervention in Psychiatry.* 2019;13(3):405–13. doi:10.1111/eip.12496
50. Lavis A, Winter R. #Online harms or benefits? An ethnographic analysis of the positives and negatives of peer-support around self-harm on social media. *Journal of Child Psychology and Psychiatry.* 2020;61(8):842–54. doi:10.1111/jcpp.13245
51. Janicke-Bowles SH, Buckley TM, Rey R, Wozniak T, Meier A, Lomanowska A. Digital Flourishing: Conceptualizing and Assessing Positive Perceptions of Mediated Social Interactions. *J Happiness Stud.* 2023 Mar 1;24(3):1013–35. doi:10.1007/s10902-023-00619-5
52. Hall JA, Steele RG, Christofferson JL, Mihailova T. Development and initial evaluation of a multidimensional digital stress scale. *Psychol Assess.* 2021 Mar;33(3):230–42. doi:10.1037/pas0000979 PubMed PMID: 33507798.
53. Lee AY, Hancock JT. Social media mindsets: a new approach to understanding social media use and psychological well-being. *Journal of Computer-Mediated Communication.* 2024 Jan 1;29(1):zmad048. doi:10.1093/jcmc/zmad048
54. Vaid SS, Kroencke L, Roshanaei M, Talaifar S, Hancock JT, Back MD, et al. Variation in social media sensitivity across people and contexts. *Sci Rep.* 2024 Mar 19;14(1):6571. doi:10.1038/s41598-024-55064-y

55. Jung K, Jang H, Kralik JD, Jeong J. Bursts and Heavy Tails in Temporal and Sequential Dynamics of Foraging Decisions. *PLoS Comput Biol*. 2014 Aug 14;10(8):e1003759. doi:10.1371/journal.pcbi.1003759 PubMed PMID: 25122498; PubMed Central PMCID: PMC4133158.
56. Siebers T, Beyens I, Valkenburg PM. The effects of fragmented and sticky smartphone use on distraction and task delay. *Mobile Media & Communication*. 2024 Jan;12(1):45–70. doi:10.1177/20501579231193941
57. Bayer JB, Anderson IA, Tokunaga RS. Building and breaking social media habits. *Current Opinion in Psychology*. 2022 Jun 1;45:101303. doi:10.1016/j.copsyc.2022.101303
58. Bayer JB, Triêu P, Ellison NB. Social Media Elements, Ecologies, and Effects. *Annual Review of Psychology*. 2020 Jan 4;71(Volume 71, 2020):471–97. doi:10.1146/annurev-psych-010419-050944
59. Orben A, Przybylski AK, Blakemore SJ, Kievit RA. Windows of developmental sensitivity to social media. *Nat Commun*. 2022 Mar 28;13(1):1649. doi:10.1038/s41467-022-29296-3
60. Tang S, Werner-Seidler A, Torok M, Mackinnon AJ, Christensen H. The relationship between screen time and mental health in young people: A systematic review of longitudinal studies. *Clinical Psychology Review*. 2021 Jun 1;86:102021. doi:10.1016/j.cpr.2021.102021
61. Parry DA, Fisher JT, Mieczkowski H, Sewall CJR, Davidson BI. Social Media and Well-being: A Methodological Perspective. *Curr Opin Psychol*. 2022 Jun;45:101285. doi:10.1016/j.copsyc.2021.11.005 PubMed PMID: 35008029; PubMed Central PMCID: PMC9167894.
62. Laibson D. Golden Eggs and Hyperbolic Discounting*. *Q J Econ*. 1997 May 1;112(2):443–78. doi:10.1162/003355397555253
63. Schulz van Endert T, Mohr PNC. Delay Discounting of Monetary and Social Media Rewards: Magnitude and Trait Effects. *Front Psychol*. 2022 Feb 11;13. doi:10.3389/fpsyg.2022.822505
64. Kleinberg J, Mullainathan S, Raghavan M. The Challenge of Understanding What Users Want: Inconsistent Preferences and Engagement Optimization [Internet]. *arXiv*; 2023 [cited 2025 Mar 25]. Available from: <http://arxiv.org/abs/2202.11776> doi:10.48550/arXiv.2202.11776
65. Hodes LN, Thomas KGF. Smartphone Screen Time: Inaccuracy of self-reports and influence of psychological and contextual factors. *Computers in Human Behavior*. 2021 Feb 1;115:106616. doi:10.1016/j.chb.2020.106616

66. Verbeij T, Pouwels JL, Beyens I, Valkenburg PM. The accuracy and validity of self-reported social media use measures among adolescents. *Computers in Human Behavior Reports*. 2021 Jan 1;3:100090. doi:10.1016/j.chbr.2021.100090
67. Ohme J, Araujo T, Boeschoten L, Freelon D, Ram N, Reeves BB, et al. Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking. *Communication Methods and Measures*. 2023;0(0):1–18. doi:10.1080/19312458.2023.2181319
68. van Driel II, Giachanou A, Pouwels JL, Boeschoten L, Beyens I, Valkenburg PM. Promises and Pitfalls of Social Media Data Donations. *Communication Methods and Measures*. 2022 Oct 2;16(4):266–82. doi:10.1080/19312458.2022.2109608
69. Cerruto F, Cirillo S, Desiato D, Gambardella SM, Polese G. Social network data analysis to highlight privacy threats in sharing data. *J Big Data*. 2022 Feb 14;9(1):19. doi:10.1186/s40537-022-00566-7
70. Boeschoten L, Ausloos J, Möller JE, Araujo T, Oberski DL. A framework for privacy preserving digital trace data collection through data donation. *Computational Communication Research*. 2022 Oct 1;4(2):388–423. doi:10.5117/CCR2022.2.002.BOES
71. The National Internet Observatory [Internet]. [cited 2026 Feb 12]. Available from: <https://nationalinternetobservatory.org/>
72. Center for Open Science. Instagram Data Access Pilot for Well-being Research [Internet]. 2025 [cited 2026 Feb 12]. Available from: <https://www.cos.io/meta>
73. Reeves B, Robinson T, Ram N. Time for the Human Screenome Project. *Nature*. 2020 Jan;577(7790):314–7. doi:10.1038/d41586-020-00032-5
74. Reeves B, Ram N, Robinson TN, Cummings JJ, Giles CL, Pan J, et al. Screenomics: A Framework to Capture and Analyze Personal Life Experiences and the Ways that Technology Shapes Them. *Hum Comput Interact*. 2021;36(2):150–201. doi:10.1080/07370024.2019.1578652 PubMed PMID: 33867652; PubMed Central PMCID: PMC8045984.
75. Massarat EAV Risa Gelles Watnick and Navid. Teens, Social Media and Technology 2022. Pew Research Center: Internet, Science & Tech [Internet]. 2022 Aug 10 [cited 2024 Mar 23]. Available from: <https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/>
76. Külling C, Waller G, Suter L, Willemse I, Bernath J, Skirgaila P, et al. JAMES : Jugend, Aktivitäten, Medien – Erhebung Schweiz [Internet]. 2022. doi:10.21256/zhaw-26216
77. Handa S, Pereira A, Holmqvist G. The Rapid Decline of Happiness: Exploring Life Satisfaction among Young People across the World. *Applied Research Quality Life*. 2023 Jun 1;18(3):1549–79. doi:10.1007/s11482-023-10153-4

78. CDC. Youth Risk Behavior Surveillance Data Summary & Trends Report: 2011-2021 [Internet]. 2021. Available from: https://www.cdc.gov/healthyyouth/data/yrbs/pdf/YRBS_Data-Summary-Trends_Report2023_508.pdf
79. UNICEF. Youth Mental Health, Study on the situation in Switzerland and Liechtenstein [Internet]. 2022 [cited 2024 May 20]. Available from: <https://www.unicef.ch/en/our-work/switzerland-liechtenstein/mental-health>
80. Marciano L, Driver CC, Schulz PJ, Camerini AL. Dynamics of adolescents' smartphone use and well-being are positive but ephemeral. *Sci Rep.* 2022 Jan 25;12(1):1. doi:10.1038/s41598-022-05291-y
81. Beyens I, Pouwels JL, van Driel II, Keijsers L, Valkenburg PM. Social Media Use and Adolescents' Well-Being: Developing a Typology of Person-Specific Effect Patterns. *Communication Research.* 2021 Dec 13;00936502211038196. doi:10.1177/00936502211038196
82. Marciano L, Vocaj E, Bekalu MA, La Tona A, Rocchi G, Viswanath K. The Use of Mobile Assessments for Monitoring Mental Health in Youth: Umbrella Review. *Journal of Medical Internet Research.* 2023;25:e45540.
83. Beyens I, Pouwels J, van Driel I, Keijsers L, Valkenburg P. Social Media Use and Adolescents' Well-Being: Developing a Typology of Person-Specific Effect Patterns. *COMMUNICATION RESEARCH.* 2021. Located at: WOS:000730999800001. doi:10.1177/00936502211038196
84. Hall JA, Xing C, Ross EM, Johnson RM. Experimentally manipulating social media abstinence: results of a four-week diary study. *Media Psychology.* 2021 Mar 4;24(2):259–75. doi:10.1080/15213269.2019.1688171
85. Marciano L, Driver CC, Schulz PJ, Camerini AL. Dynamics of adolescents' smartphone use and well-being are positive but ephemeral. *Sci Rep.* 2022 Jan 25;12(1):1. doi:10.1038/s41598-022-05291-y
86. Domoff SE, Banga CA, Borgen AL, Foley RP, Robinson C, Avery K, et al. Use of passive sensing to quantify adolescent mobile device usage: Feasibility, acceptability, and preliminary validation of the eMoodie application. *Human Behavior and Emerging Technologies.* 2021;3(1):63–74. doi:10.1002/hbe2.247
87. Ohme J, Araujo T, de Vreese CH, Piotrowski JT. Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function. *Mobile Media & Communication.* 2021 May 1;9(2):293–313. doi:10.1177/2050157920959106
88. Marciano L, Camerini AL. Duration, frequency, and time distortion: Which is the best predictor of problematic smartphone use in adolescents? A trace data study. *PLoS One.* 2022;17(2):e0263815. doi:10.1371/journal.pone.0263815 PubMed PMID: 35180248; PubMed Central PMCID: PMC8856513.

89. Parry DA, Davidson BI, Sewall CJR, Fisher JT, Mieczkowski H, Quintana DS. A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nat Hum Behav*. 2021 May 17;1–13. doi:10.1038/s41562-021-01117-5
90. Directorate-General for Communication. The EU's Digital Services Act [Internet]. 2022 [cited 2024 Oct 24]. Available from: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en
91. Scott Bomboy. National Constitution Center – constitutioncenter.org [Internet]. [cited 2024 Oct 24]. The Supreme Court considers state laws regulating social media giants | Constitution Center. Available from: <https://constitutioncenter.org/blog/the-supreme-court-considers-state-laws-regulating-social-media-giants>
92. Sultan M, Scholz C, van den Bos W. Leaving traces behind: Using social media digital trace data to study adolescent wellbeing. *Computers in Human Behavior Reports*. 2023 May 1;10:100281. doi:10.1016/j.chbr.2023.100281
93. Davidson BI, Wischerath D, Racek D, Parry DA, Godwin E, Hinds J, et al. Platform-controlled social media APIs threaten Open Science [Internet]. OSF; 2023 [cited 2024 Mar 23]. Available from: <https://osf.io/ps32z> doi:10.31234/osf.io/ps32z
94. Murtfeldt R, Alterman N, Kahveci I, West JD. RIP Twitter API: A eulogy to its vast research contributions [Internet]. arXiv; 2024 [cited 2024 May 20]. Available from: <http://arxiv.org/abs/2404.07340> doi:10.48550/arXiv.2404.07340
95. Wright JC. Stakeholder Management in Change Initiatives: Reddit Changes Its API Pricing [Internet]. London; 2024 [cited 2024 May 20]. Available from: <https://sk.sagepub.com/cases/stakeholder-management-in-change-initiatives-reddit-changes?token=71f3550c-c5fb-4203-9b31-b23df4026441788cf34ed398d7dd60c9369addd8c0a587d1ad7e0a85bb42bfb5191a7ec03a54> doi:10.4135/9781071941379
96. Lind MN, Kahn LE, Crowley R, Reed W, Wicks G, Allen NB. Reintroducing the Effortless Assessment Research System (EARS). *JMIR Ment Health*. 2023 Apr 26;10:e38920. doi:10.2196/38920 PubMed PMID: 37099361; PubMed Central PMCID: PMC10173035.
97. Stier S, Breuer J, Siegers P, Thorson K. Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*. 2020;38(5):503–16. doi:10.1177/0894439319843669
98. Ram N, Haber N, Robinson TN, Reeves B. Binding the Person-Specific Approach to Modern AI in the Human Screenome Project: Moving past Generalizability to Transferability. *Multivariate Behavioral Research*. 2023;0(0):1–9. doi:10.1080/00273171.2023.2229305

99. Liu, Yuning, Klassen, Geoff, Mee, Jenna, Pointer, Justin, Marvi, Baloch, Marciano, Laura, et al. User-donated screenshots analysis: feasibility of a new approach to collect objective social media app usage in adolescents. In: SBP-BRiMS 2024. Pittsburgh, PA; 2024.
100. Leyland AH, Groenewegen PP. What Is Multilevel Modelling? In: Multilevel Modelling for Public Health and Health Services Research: Health in Context [Internet] [Internet]. Springer; 2020 [cited 2024 May 20]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK565712/> doi:10.1007/978-3-030-34801-4_3
101. Orben A, Dienlin T, Przybylski AK. Social media's enduring effect on adolescent life satisfaction. *Proc Natl Acad Sci U S A*. 2019 May 21;116(21):10226–8. doi:10.1073/pnas.1902058116 PubMed PMID: 31061122; PubMed Central PMCID: PMC6534991.
102. Hall JA, Liu D. Social media use, social displacement, and well-being. *Current Opinion in Psychology*. 2022 Aug 1;46:101339. doi:10.1016/j.copsyc.2022.101339
103. Politte-Corn M, Dickey L, Abitante G, Pegg S, Bean CAL, Kujawa A. Social Media Use as a Predictor of Positive and Negative Affect: An Ecological Momentary Assessment Study of Adolescents with and without Clinical Depression. *Res Child Adolesc Psychopathol*. 2024 Feb 20. doi:10.1007/s10802-024-01177-x
104. Johannes N, Masur P, Vuorre M, Przybylski A. How should we investigate variation in the relation between social media and well-being [Internet]. 2021 [cited 2024 May 20]. Available from: <https://doi.org/10.31234/osf.io/xahbg> doi:10.31234/osf.io/xahbg
105. Verduyn P, Lee DS, Park J, Shablack H, Orvell A, Bayer J, et al. Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence. *Journal of Experimental Psychology: General*. 2015 Apr;144(2):480–8. Located at: 2015-08049-001. doi:10.1037/xge0000057
106. Wu Y, An Z, Lin Y, Zhang J, Jing B, Peng K. Social media use, uncertainty, and negative affect in times of pandemic crisis. *Digit Health*. 2023 Jun 12;9:20552076231181227. doi:10.1177/20552076231181227 PubMed PMID: 37334319; PubMed Central PMCID: PMC10272698.
107. Firth J, Torous J, Stubbs B, Firth JA, Steiner GZ, Smith L, et al. The “online brain”: how the Internet may be changing our cognition. *World Psychiatry*. 2019 Jun;18(2):119–29. doi:10.1002/wps.20617 PubMed PMID: 31059635; PubMed Central PMCID: PMC6502424.
108. Bayer JB, Triêu P, Ellison NB. Social Media Elements, Ecologies, and Effects. *Annual Review of Psychology*. 2020 Jan 4;71(Volume 71, 2020):471–97. doi:10.1146/annurev-psych-010419-050944
109. Masciantonio A, Bourguignon D, Bouchat P, Balty M, Rimé B. Don't put all social network sites in one basket: Facebook, Instagram, Twitter, TikTok, and their relations

- with well-being during the COVID-19 pandemic. *PLOS ONE*. 2021 Mar 11;16(3):e0248384. doi:10.1371/journal.pone.0248384
110. Faelens L, Hoorelbeke K, Soenens B, Van Gaeveren K, De Marez L, De Raedt R, et al. Social media use and well-being: A prospective experience-sampling study. *Computers in Human Behavior*. 2021 Jan 1;114:106510. doi:10.1016/j.chb.2020.106510
 111. Matthes J, Karsay K, Schmuck D, Stevic A. “Too much to handle”: Impact of mobile social networking sites on information overload, depressive symptoms, and well-being. *Computers in Human Behavior*. 2020 Apr 1;105:106217. doi:10.1016/j.chb.2019.106217
 112. Masciantonio A, Résibois M, Bouchat P, Bourguignon D. Social Network Sites and Well-Being: Is it Only a Matter of Content? *International Review of Social Psychology*. 2023 May 5;36(1). doi:10.5334/irsp.736
 113. Bowman ND, Rieger D, Tammy Lin JH. Social video gaming and well-being. *Current Opinion in Psychology*. 2022 Jun 1;45:101316. doi:10.1016/j.copsyc.2022.101316
 114. Vuorre M, Johannes N, Magnusson K, Przybylski AK. Time spent playing video games is unlikely to impact well-being. *Royal Society Open Science*. 2022 Jul 27;9(7):220411. doi:10.1098/rsos.220411
 115. Reer F, Quandt T. Digital Games and Well-Being: An Overview. In: Kowert R, editor. *Video Games and Well-being: Press Start* [Internet]. Cham: Springer International Publishing; 2020 [cited 2024 Oct 24]. p. 1–21. Available from: https://doi.org/10.1007/978-3-030-32770-5_1 doi:10.1007/978-3-030-32770-5_1
 116. Aziz N, Nordin MJ, Abdulkadir SJ, Salih MMM. Digital Addiction: Systematic Review of Computer Game Addiction Impact on Adolescent Physical Health. *Electronics*. 2021 Jan;10(9):9. doi:10.3390/electronics10090996
 117. Eni Purwaningsih -, Ira Nurmala -. The Impact of Online Game Addiction on Adolescent Mental Health: A Systematic Review and Meta-analysis. *Open Access Macedonian Journal of Medical Sciences (OAMJMS)*. 2021;9(F):F.
 118. Nilsson A, Rosendahl I, Jayaram-Lindström N. Gaming and social media use among adolescents in the midst of the COVID-19 pandemic. *Nordic Studies on Alcohol and Drugs*. 2022 Aug 1;39(4):347–61. doi:10.1177/14550725221074997
 119. Marciano L, Viswanath K. Social media use and adolescents’ well-being: A note on flourishing. *Front Psychol*. 2023 Apr 6;14. doi:10.3389/fpsyg.2023.1092109
 120. Ryff CD, Dienberg Love G, Urry HL, Muller D, Rosenkranz MA, Friedman EM, et al. Psychological well-being and ill-being: do they have distinct or mirrored biological correlates? *Psychother Psychosom*. 2006;75(2):85–95. doi:10.1159/000090892 PubMed PMID: 16508343.

121. Janicke-Bowles SH, Buckley TM, Rey R, Wozniak T, Meier A, Lomanowska A. Digital Flourishing: Conceptualizing and Assessing Positive Perceptions of Mediated Social Interactions. *J Happiness Stud.* 2023 Mar 1;24(3):1013–35. doi:10.1007/s10902-023-00619-5
122. VanderWeele TJ, McNeely E, Koh HK. Reimagining Health—Flourishing. *JAMA.* 2019 May 7;321(17):1667–8. doi:10.1001/jama.2019.3035
123. Ryff CD, Boylan JM, Kirsch JA. Eudaimonic and hedonic well-being: An integrative perspective with linkages to sociodemographic factors and health. In: *Measuring well-being: Interdisciplinary perspectives from the social sciences and the humanities.* New York, NY, US: Oxford University Press; 2021. p. 92–135. doi:10.1093/oso/9780197512531.003.0005
124. Bates D, Maechler M, Bolker [aut B, cre, Walker S, Christensen RHB, et al. lme4: Linear Mixed-Effects Models using “Eigen” and S4 [Internet]. 2024 [cited 2024 Jul 22]. Available from: <https://cran.r-project.org/web/packages/lme4/index.html>
125. Enders CK, Keller BT, Woller MP. A simple Monte Carlo method for estimating power in multilevel designs. *Psychol Methods.* 2023 Nov 13. doi:10.1037/met0000614 PubMed PMID: 37956081.
126. Silvianetri S, Irman I, Zubaidah Z, Yeni P, Syafitri R. Student Preferences Watching Youtube and Its Implications for Wellbeing Counseling. In. 2022 [cited 2024 May 21]. Available from: <https://eudl.eu/doi/10.4108/eai.11-10-2021.2319487>
127. Balcombe L, De Leo D. The Impact of YouTube on Loneliness and Mental Health. *Informatics.* 2023 Jun;10(2):2. doi:10.3390/informatics10020039
128. Mackson SB, Brochu PM, Schneider BA. Instagram: Friend or foe? The application’s association with psychological well-being. *New Media & Society.* 2019 Oct 1;21(10):2160–82. doi:10.1177/1461444819840021
129. Roberts JA, David ME. Instagram and TikTok Flow States and Their Association with Psychological Well-Being. *Cyberpsychology, Behavior, and Social Networking.* 2023 Feb;26(2):80–9. doi:10.1089/cyber.2022.0117
130. Bestvater S. Acknowledgments. Pew Research Center [Internet]. 2024 Feb 22 [cited 2024 Oct 24]. Available from: <https://www.pewresearch.org/internet/2024/02/22/acknowledgments-38-3/>
131. Marciano L, Ostroumova M, Schulz PJ, Camerini AL. Digital Media Use and Adolescents’ Mental Health During the Covid-19 Pandemic: A Systematic Review and Meta-Analysis. *Front Public Health.* 2021;9:793868. doi:10.3389/fpubh.2021.793868 PubMed PMID: 35186872; PubMed Central PMCID: PMC8848548.

132. Karim HA, Kwan SG. WOMEN, STRESS AND COPING WITH TIKTOK. MANU Jurnal Pusat Penataran Ilmu dan Bahasa. 2024 Jun 28;35(1):1. doi:10.51200/manu.v35i1.5160
133. Bano S, Cisheng W, Khan AN, Khan NA. WhatsApp use and student's psychological well-being: Role of social capital and social integration. Children and Youth Services Review. 2019 Aug 1;103:200–8. doi:10.1016/j.childyouth.2019.06.002
134. Wambua TN, Muriithi DJK, Ikpe DRO. WhatsApp Usage and Psychological Well-Being of Students in Mixed Secondary Schools in Masinga Sub County, Machakos County, Kenya [Internet]. 2022 Dec [cited 2024 May 20]. Available from: <https://ir.gretnsauniversity.ac.ke/xmlui/handle/20.500.12736/3775>
135. Blabst N, Diefenbach S. WhatsApp and Wellbeing: A study on WhatsApp usage, communication quality and stress. In. BCS Learning & Development; 2017 [cited 2024 May 20]. Available from: <https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/HCI2017.85> doi:10.14236/ewic/HCI2017.85
136. Vannucci A, McCauley Ohannessian C. Social Media Use Subgroups Differentially Predict Psychosocial Well-Being During Early Adolescence. J Youth Adolescence. 2019 Aug 1;48(8):1469–93. doi:10.1007/s10964-019-01060-9
137. Dunn TR, Langlais MR. “Oh, Snap!”: A Mixed-Methods Approach to Analyzing the Dark Side of Snapchat. The Journal of Social Media in Society. 2020 Dec 18;9(2):2.
138. Li Y, Xu Z, Hao Y, Xiao P, Liu J. Psychosocial Impacts of Mobile Game on K12 Students and Trend Exploration for Future Educational Mobile Games. Front Educ. 2022 Apr 29;7. doi:10.3389/educ.2022.843090
139. Patti Valkenburg, Amber van der Wal, Ine Beyens. De invloed van social media op de mentale gezondheid van jongeren [Internet]. UNICEF; 2024. Available from: https://pure.uva.nl/ws/files/175927914/Schermgeluk_en_schermverdriet.pdf
140. O'Reilly M, Dogra N, Hughes J, Reilly P, George R, Whiteman N. Potential of social media in promoting mental health in adolescents. Health Promotion International. 2019 Oct 1;34(5):981–91. doi:10.1093/heapro/day056
141. Beyens I, Pouwels JL, van Driel II, Keijsers L, Valkenburg PM. The effect of social media on well-being differs from adolescent to adolescent. Sci Rep. 2020 Jul 1;10(1):10763. doi:10.1038/s41598-020-67727-7
142. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The Extent and Consequences of P-Hacking in Science. PLOS Biology. 2015 Mar 13;13(3):e1002106. doi:10.1371/journal.pbio.1002106
143. Wen CKF, Schneider S, Stone AA, Spruijt-Metz D. Compliance With Mobile Ecological Momentary Assessment Protocols in Children and Adolescents: A Systematic Review and

- Meta-Analysis. *J Med Internet Res*. 2017 Apr 26;19(4):e132. doi:10.2196/jmir.6641 PubMed PMID: 28446418; PubMed Central PMCID: PMC5425774.
144. Wrzus C, Neubauer AB. Ecological Momentary Assessment: A Meta-Analysis on Designs, Samples, and Compliance Across Research Fields. *Assessment*. 2023 Apr;30(3):825–46. doi:10.1177/10731911211067538 PubMed PMID: 35016567; PubMed Central PMCID: PMC9999286.
 145. Park JG and E. Americans' Social Media Use 2025. Pew Research Center [Internet]. 2025 Nov 20 [cited 2025 Nov 21]. Available from: <https://www.pewresearch.org/internet/2025/11/20/americans-social-media-use-2025/>
 146. Meta Reports Third Quarter 2025 Results [Internet]. [cited 2026 Jan 4]. Available from: <https://investor.atmeta.com/investor-news/press-release-details/2025/Meta-Reports-Third-Quarter-2025-Results/default.aspx>
 147. blog.youtube [Internet]. [cited 2026 Jan 4]. YouTube for Press. Available from: <https://blog.youtube/press/>
 148. Braghieri L, Levy R, Makarin A. Social Media and Mental Health. *American Economic Review*. 2022 Nov;112(11):3660–93. doi:10.1257/aer.20211218
 149. Locke EA, Latham GP. New Directions in Goal-Setting Theory. *Curr Dir Psychol Sci*. 2006 Oct 1;15(5):265–8. doi:10.1111/j.1467-8721.2006.00449.x
 150. Dijksterhuis A, Aarts H. Goals, Attention, and (Un)Consciousness. *Annual Review of Psychology*. 2010 Jan 1;61(Volume 61, 2010):467–90. doi:10.1146/annurev.psych.093008.100445
 151. Kruglanski AW, Shah JY, Fishbach A, Friedman R, Chun WY, Sleeth-Keppler D. A theory of goal systems. In: *The Motivated Mind*. Routledge; 2018.
 152. Klein HJ, Wesson MJ, Hollenbeck JR, Alge BJ, Klein HJ, Wesson MJ, et al. Goal commitment and the goal-setting process: conceptual clarification and empirical synthesis. *Journal of Applied Psychology*. 1999 Dec 1;84(6):885–96. Located at: 12128415. doi:10.1037/0021-9010.84.6.885
 153. Gutentag T, Kalokerinos EK, Millgram Y, Garrett PM, Sobel R, Tamir M. Motivational Intensity in Emotion Regulation. *Pers Soc Psychol Bull*. 2026 Jan 1;52(1):212–26. doi:10.1177/01461672241273273
 154. Millgram Y, Nock MK, Bailey DD, Goldenberg A. Knowledge About the Source of Emotion Predicts Emotion-Regulation Attempts, Strategies, and Perceived Emotion-Regulation Success. *Psychol Sci*. 2023 Nov;34(11):1244–55. doi:10.1177/09567976231199440 PubMed PMID: 37796082.

155. Frances Haugen [producer]. Cambridge, MA: FBarchive [distributor] [Internet]. 2023 [cited 2026 Jan 10]. FBarchive Document odoc000805w35 (Green Edition). Available from: <https://fbarchive.org/user/doc/odoc000805w35>
156. Lubin N, Liu Y, Yarnell A, Austin SB, Ward ZJ, Iyer R, et al. Social Media Harm Abatement: Mechanisms for Transparent Public Health Assessment [Internet]. arXiv; 2025 [cited 2025 Mar 14]. Available from: <http://arxiv.org/abs/2503.10458> doi:10.48550/arXiv.2503.10458
157. Lai CH. Motivations, Usage, and Perceived Social Networks Within and Beyond Social Media. *J Comput Mediat Commun*. 2019 May 1;24(3):126–45. doi:10.1093/jcmc/zmz004
158. Throuvala MA, Griffiths MD, Rennoldson M, Kuss DJ. Motivational processes and dysfunctional mechanisms of social media use among adolescents: A qualitative focus group study. *Computers in Human Behavior*. 2019 Apr 1;93:164–75. doi:10.1016/j.chb.2018.12.012
159. Lee AY, Ellison NB, Hancock JT. To use or be used? The role of agency in social media use and well-being. *Front Comput Sci*. 2023 Apr 3;5. doi:10.3389/fcomp.2023.1123323
160. Douglas P, Coetzee B. ResearchGate [Internet]. 2024 [cited 2025 Apr 8]. Do Mindsets Really Matter? A Closer Look at Social Media Perceptions and Well-Being | Request PDF. Available from: https://www.researchgate.net/publication/389619842_Do_Mindsets_Really_Matter_A_Closer_Look_at_Social_Media_Perceptions_and_Well-Being
161. Papacharissi Z, Mendelson A. Toward a new(er) sociability: Uses, gratifications and social capital on Facebook. *Media Perspectives for the 21st Century*. 2011 Jan 1;212–30. doi:10.4324/9780203834077
162. Smock AD, Ellison NB, Lampe C, Wohn DY. Facebook as a toolkit: A uses and gratification approach to unbundling feature use. *Computers in Human Behavior*. 2011 Nov 1;27(6):2322–9. doi:10.1016/j.chb.2011.07.011
163. Lee EJ, Kim YW. How social is Twitter use? Affiliative tendency and communication competence as predictors. *Computers in Human Behavior*. 2014 Oct 1;39:296–305. doi:10.1016/j.chb.2014.07.034
164. Rigby CS, Ryan RM. Time Well-Spent?: Motivation for Entertainment Media and Its Eudaimonic Aspects Through the Lens of Self-Determination Theory. In: *The Routledge Handbook of Media Use and Well-Being*. Routledge; 2016.
165. Wolfers LN, Utz S. Social media use, stress, and coping. *Current Opinion in Psychology*. 2022 Jun 1;45:101305. doi:10.1016/j.copsyc.2022.101305
166. Bengtsson S, Johansson S. The Meanings of Social Media Use in Everyday Life: Filling Empty Slots, Everyday Transformations, and Mood Management. *Social Media + Society*. 2022 Oct 1;8(4):20563051221130292. doi:10.1177/20563051221130292

167. Vishwanath A. Habitual Facebook Use and its Impact on Getting Deceived on Social Media. *J Comput Mediat Commun*. 2015 Jan 1;20(1):83–98. doi:10.1111/jcc4.12100
168. Anderson IA, Wood W. Habits and the electronic herd: The psychology behind social media's successes and failures. *Consumer Psychology Review*. 2021;4(1):83–99. doi:10.1002/arcp.1063
169. Maza MT, Fox KA, Kwon SJ, Flannery JE, Lindquist KA, Prinstein MJ, et al. Association of Habitual Checking Behaviors on Social Media With Longitudinal Functional Brain Development. *JAMA Pediatr*. 2023 Feb 1;177(2):160–7. doi:10.1001/jamapediatrics.2022.4924
170. Ouellette JA, Wood W. Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin*. 1998 Jul;124(1):54–74. Located at: 1998-04232-003. doi:10.1037/0033-2909.124.1.54
171. Wood W, Mazar A, Neal DT. Habits and Goals in Human Behavior: Separate but Interacting Systems. *Perspect Psychol Sci*. 2022 Mar 1;17(2):590–605. doi:10.1177/1745691621994226
172. Lin KY, Lu HP. Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior*. 2011 May 1;Group Awareness in CSCL Environments27(3):1152–61. doi:10.1016/j.chb.2010.12.009
173. Seijts GH, Latham GP. The Construct of Goal Commitment: Measurement And Relationships with Task Performance. In: Goffin RD, Helmes E, editors. *Problems and Solutions in Human Assessment: Honoring Douglas N. Jackson at Seventy* [Internet]. Boston, MA: Springer US; 2000 [cited 2026 Jan 8]. p. 315–32. Available from: https://doi.org/10.1007/978-1-4615-4397-8_14 doi:10.1007/978-1-4615-4397-8_14
174. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *Am Psychol*. 2002 Sep;57(9):705–17. doi:10.1037//0003-066x.57.9.705 PubMed PMID: 12237980.
175. Wigfield A, Eccles JS. Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology*. 2000 Jan 1;25(1):68–81. doi:10.1006/ceps.1999.1015
176. Baker RK, White KM. Predicting adolescents' use of social networking sites from an extended theory of planned behaviour perspective. *Computers in Human Behavior*. 2010 Nov 1;Online Interactivity: Role of Technology in Behavior Change26(6):1591–7. doi:10.1016/j.chb.2010.06.006
177. Pelling EL, White KM. The Theory of Planned Behavior Applied to Young People's Use of Social Networking Web Sites. *CyberPsychology & Behavior*. 2009 Dec;12(6):755–9. doi:10.1089/cpb.2009.0109

178. Van de Castele M, Soenens B, Ponnet K, Perneel S, Flamant N, Vansteenkiste M. Unraveling the Role of Social Media on Adolescents' Daily Goals and Affect: The Interplay Between Basic Psychological Needs and Screen Time [Internet]. [cited 2026 Jan 11]. Available from: <https://dx.doi.org/10.1093/iwc/iwad055>
179. Gilbert A, Baumgartner SE, Reinecke L. Situational boundary conditions of digital stress: Goal conflict and autonomy frustration make smartphone use more stressful. *Mobile Media & Communication*. 2023 Sep 1;11(3):435–58. doi:10.1177/20501579221138017
180. Duckworth AL, Peterson C, Matthews MD, Kelly DR. Grit: perseverance and passion for long-term goals. *J Pers Soc Psychol*. 2007 Jun;92(6):1087–101. doi:10.1037/0022-3514.92.6.1087 PubMed PMID: 17547490.
181. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin*. 2006 Mar 1;132(2):249–68. Located at: 2006-03023-004. doi:10.1037/0033-2909.132.2.249
182. Turel O, Brevers D, Bechara A. Time distortion when users at-risk for social media addiction engage in non-social media tasks. *Journal of Psychiatric Research*. 2018 Feb 1;97:84–8. doi:10.1016/j.jpsychires.2017.11.014
183. Scott H, Biello SM, Woods HC. Identifying drivers for bedtime social media use despite sleep costs: The adolescent perspective. *Sleep Health*. 2019 Dec 1;5(6):539–45. doi:10.1016/j.sleh.2019.07.006
184. Adeoye AO, Jegede AS. Use of digital social media at night, awareness of its effect on sleep, potential risks and factors driving the behaviour. *J Public Health (Berl)*. 2024 Jan 1;32(1):103–9. doi:10.1007/s10389-022-01794-0
185. Gollwitzer P, Bargh J. Automaticity in Goal Pursuit. In. 2005 [cited 2026 Jan 11]. Available from: <https://www.semanticscholar.org/paper/Automaticity-in-Goal-Pursuit-Gollwitzer-Bargh/39b671234a71b6f3ef11e9f1888a7e1124251b22>
186. Custers R, Vermeent S, Aarts H. Does Goal Pursuit Require Conscious Awareness? In: Ryan RM, editor. *The Oxford Handbook of Human Motivation* [Internet]. Oxford University Press; 2019 [cited 2026 Jan 11]. p. 0. Available from: <https://doi.org/10.1093/oxfordhb/9780190666453.013.15> doi:10.1093/oxfordhb/9780190666453.013.15
187. Bongers KCA, Dijksterhuis A, Spears R. Self-esteem regulation after success and failure to attain unconsciously activated goals. *Journal of Experimental Social Psychology*. 2009 May 1;45(3):468–77. doi:10.1016/j.jesp.2008.12.007
188. Shah JY. The Automatic Pursuit and Management of Goals. *Curr Dir Psychol Sci*. 2005 Feb 1;14(1):10–3. doi:10.1111/j.0963-7214.2005.00325.x
189. Millgram Y, Nock M, Bailey D, Goldenberg A. Knowledge about the source of emotion predicts emotion regulation attempts, strategies and perceived emotion regulation success

- [Internet]. 2022 [cited 2024 Apr 5]. Available from: <https://doi.org/10.31234/osf.io/p2bmq>
doi:10.31234/osf.io/p2bmq
190. Bestvater S. How U.S. Adults Use TikTok. Pew Research Center [Internet]. 2024 Feb 22 [cited 2024 Oct 24]. Available from:
<https://www.pewresearch.org/internet/2024/02/22/how-u-s-adults-use-tiktok/>
 191. Andreassen CS, Torsheim T, Brunborg GS, Pallesen S. Development of a Facebook Addiction Scale. *Psychol Rep.* 2012 Apr;110(2):501–17.
doi:10.2466/02.09.18.PR0.110.2.501-517 PubMed PMID: 22662404.
 192. Ltd BPG. The 2-item Generalized Anxiety Disorder scale had high sensitivity and specificity for detecting GAD in primary care. *BMJ Evidence-Based Medicine.* 2007 Oct 1;12(5):149–149. doi:10.1136/ebm.12.5.149 PubMed PMID: 17909240.
 193. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care.* 2003 Nov;41(11):1284–92.
doi:10.1097/01.MLR.0000093487.78664.3C PubMed PMID: 14583691.
 194. Cacioppo JT, Petty RE, Kao CF. The efficient assessment of need for cognition. *J Pers Assess.* 1984 Jun;48(3):306–7. doi:10.1207/s15327752jpa4803_13 PubMed PMID: 16367530.
 195. Wong SL, McLellan R. Adolescents’ motivation to use social network sites from a psychological needs perspective. *Journal of Adolescence.* 2024;96(2):305–21.
doi:10.1002/jad.12276
 196. Marcus EB. How to Build Mindful Social Media Habits in the Age of TikTok—for You and Your Kids. *Mindful* [Internet]. 2025 Jan 17 [cited 2026 Feb 10]. Available from:
<https://www.mindful.org/building-compassionate-social-media-habits-for-you-and-your-kids/>
 197. Fracchia G. Redesign and Nudging Strategies for Digital Wellbeing on YouTube [laurea] [Internet]. Politecnico di Torino; 2023 [cited 2024 Apr 19]. Available from:
<https://webthesis.biblio.polito.it/29497/>
 198. Harambam J, Bountouridis D, Makhortykh M, van Hoboken J. Designing for the better by taking users into account: a qualitative evaluation of user control mechanisms in (news) recommender systems. In: *Proceedings of the 13th ACM Conference on Recommender Systems* [Internet]. New York, NY, USA: Association for Computing Machinery; 2019 [cited 2024 Jul 2]. p. 69–77. (RecSys ’19). Available from:
<https://doi.org/10.1145/3298689.3347014> doi:10.1145/3298689.3347014
 199. Lyngs U, Lukoff K, Slovak P, Binns R, Slack A, Inzlicht M, et al. Self-Control in Cyberspace: Applying Dual Systems Theory to a Review of Digital Self-Control Tools. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* [Internet]. New York, NY, USA: Association for Computing Machinery; 2019 [cited 2024

- Jul 1]. p. 1–18. (CHI '19). Available from: <https://doi.org/10.1145/3290605.3300361>
doi:10.1145/3290605.3300361
200. Caraban A, Karapanos E, Gonçalves D, Campos P. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2019 [cited 2024 Apr 21]. p. 1–15. (CHI '19). Available from: <https://dl.acm.org/doi/10.1145/3290605.3300733>
doi:10.1145/3290605.3300733
 201. Grüning DJ, Riedel F, Lorenz-Spreen P. Directing smartphone use through the self-nudge app one sec. Proceedings of the National Academy of Sciences. 2023 Feb 21;120(8):e2213114120. doi:10.1073/pnas.2213114120
 202. Cut your screen time in half | one sec [Internet]. 2025 [cited 2026 Feb 10]. Available from: <https://one-sec.app/>
 203. Google Play. Digital Wellbeing - Apps on Google Play [Internet]. [cited 2026 Feb 10]. Available from:
https://play.google.com/store/apps/details?id=com.google.android.apps.wellbeing&hl=en_US
 204. Schnabel T, Bennett PN, Dumais ST, Joachims T. Short-Term Satisfaction and Long-Term Coverage: Understanding How Users Tolerate Algorithmic Exploration. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining [Internet]. New York, NY, USA: Association for Computing Machinery; 2018 [cited 2026 Feb 10]. p. 513–21. (WSDM '18). Available from:
<https://doi.org/10.1145/3159652.3159700> doi:10.1145/3159652.3159700
 205. Reinecke L, Gilbert A, Eden A. Self-regulation as a key boundary condition in the relationship between social media use and well-being. Current Opinion in Psychology. 2022 Jun 1;45:101296. doi:10.1016/j.copsyc.2021.12.008
 206. Arvind Narayanan, Sayash Kapoor. AI Snake Oil | Princeton University Press [Internet]. 2024 [cited 2026 Feb 10]. Available from:
<https://press.princeton.edu/books/hardcover/9780691249131/ai-snake-oil>
 207. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology. 2003;88(5):879–903. doi:10.1037/0021-9010.88.5.879
 208. Mahalingham T, McEvoy PM, Clarke PJF. Assessing the validity of self-report social media use: Evidence of No relationship with objective smartphone use. Computers in Human Behavior. 2023 Mar 1;140:107567. doi:10.1016/j.chb.2022.107567
 209. Mojescik KM, Berens SC, De Luca F, Ritchey M, Bird CM. The Relationship Between Subjective Memory Experience and Objective Memory Performance Remains Stable

- Across the Lifespan. Mathôt S, editor. Collabra: Psychology. 2024 Apr 29;10(1):116195. doi:10.1525/collabra.116195
210. Chidharom M, Bonnefond A, Vogel EK, Rosenberg MD. Objective markers of sustained attention fluctuate independently of mind-wandering reports. *Psychon Bull Rev.* 2025 Aug 1;32(4):1689–99. doi:10.3758/s13423-025-02640-6
 211. Gjorup ALT, Maia A, Schmidt GJ, Marques AC, Junior JCT, Trotto EF, et al. Discrepancy between self-reported and performance-based attention. *Journal of Affective Disorders.* 2026 Apr 15;399:121098. doi:10.1016/j.jad.2025.121098
 212. Self-Perception Theory. In: *Advances in Experimental Social Psychology* [Internet]. Academic Press; 1972 [cited 2026 Feb 11]. p. 1–62. Available from: <https://www.sciencedirect.com/science/chapter/bookseries/abs/pii/S0065260108600246> doi:10.1016/S0065-2601(08)60024-6
 213. Carrière TC, Boeschoten L, Struminskaya B, Janssen HL, de Schipper NC, Araujo T. Best practices for studies using digital data donation. *Qual Quant.* 2024 Oct 8. doi:10.1007/s11135-024-01983-x
 214. Thorbecke L, Agarwal AK, Guntuku SC. Pragmatic lessons for harnessing social media for health research through digital data donation post-API access. *npj Digit Public Health.* 2026 Jan 21;1(1):2. doi:10.1038/s44482-025-00002-6
 215. Manoel Horta Ribeiro, Andrew Guess. Princeton’s Digital Data Donation Initiative | Data-Driven Social Science [Internet]. 2025 [cited 2026 Feb 11]. Available from: <https://ddss.princeton.edu/research/featured-projects/princeton%E2%80%99s-digital-data-donation-initiative>
 216. JISC [Internet]. 2022 [cited 2025 Feb 24]. Digital wellbeing. Available from: <https://digitalcapability.jisc.ac.uk/what-is-digital-capability/digital-wellbeing/>
 217. Roffarello AM, De Russis L. Achieving Digital Wellbeing Through Digital Self-control Tools: A Systematic Review and Meta-analysis. *ACM Trans Comput-Hum Interact.* 2023 Sep 12;30(4):53:1-53:66. doi:10.1145/3571810
 218. Zhang MR, Lukoff K, Rao R, Baughan A, Hiniker A. Monitoring Screen Time or Redesigning It? Two Approaches to Supporting Intentional Social Media Use. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* [Internet]. New York, NY, USA: Association for Computing Machinery; 2022 [cited 2024 Jul 1]. p. 1–19. (CHI ’22). Available from: <https://doi.org/10.1145/3491102.3517722> doi:10.1145/3491102.3517722
 219. Lukoff K, Lyngs U, Shirokova K, Rao R, Tian L, Zade H, et al. SwitchTube: A Proof-of-Concept System Introducing “Adaptable Commitment Interfaces” as a Tool for Digital Wellbeing. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* [Internet]. New York, NY, USA: Association for Computing Machinery; 2023

- [cited 2024 Jul 1]. p. 1–22. (CHI '23). Available from:
<https://doi.org/10.1145/3544548.3580703> doi:10.1145/3544548.3580703
220. Purohit A, Holzer A. Unhooked by Design: Scrolling Mindfully on Social Media by Automating Digital Nudges. In. 2021.
 221. Children and Screens. UK Age-Appropriate Design Code: Impact assessment [Internet]. 2024. Report Institute of Digital Media and Child Development. Available from:
<https://www.childrenandscreens.org/wp-content/uploads/2024/03/Children-and-Screens-UK-AADC-Impact-Assessment.pdf>
 222. Steve Wood. Impact of regulation on children’s digital lives [Internet]. 2024. Available from:
https://eprints.lse.ac.uk/123522/1/Impact_of_regulation_on_children_DFC_Research_report_May_2024.pdf
 223. Ravi Iyer, Yuning Liu. Social media harm taxonomy [Internet]. 2024. Available from:
<https://docs.google.com/spreadsheets/d/1pqaBBgf6ZIdLyGHRuapgIWeCJ6NUw6KIIrvQmMpZeZ8/edit?usp=sharing>
 224. Ravi Iyer. Taxonomy - Mapping Consumer Harm to Specific Design Elements [Internet]. 2025. Available from: https://docs.google.com/spreadsheets/d/10-ojejxc6a33-1fS0O-prVZgu02_GjNHjm89vNSxQ5s/edit?usp=sharing
 225. Suter V, Meckel M. Using GPT-4 for Text Analysis: Insights from English and German Language News Classification Tasks. Workshop Proceedings of the 18th International AAAI Conference on Web and Social Media. 2024 Jun 1;2024:31. doi:10.36190/2024.31
 226. Gao J, Guo Y, Lim G, Zhang T, Zhang Z, Li TJJ, et al. CollabCoder: A Lower-barrier, Rigorous Workflow for Inductive Collaborative Qualitative Analysis with Large Language Models. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2024 [cited 2025 Feb 18]. p. 1–29. (CHI '24). Available from:
<https://dl.acm.org/doi/10.1145/3613904.3642002> doi:10.1145/3613904.3642002
 227. NW 1615 L. St, Washington S 800, Inquiries D 20036 U 419 4300 | M 857 8562 | F 419 4372 | M. Social Media Fact Sheet. Pew Research Center: Internet, Science & Tech [Internet]. 2024 [cited 2024 Apr 15]. Available from:
<https://www.pewresearch.org/internet/fact-sheet/social-media/>
 228. Reimers N, Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks [Internet]. arXiv; 2019 [cited 2026 Jan 5]. Available from:
<http://arxiv.org/abs/1908.10084> doi:10.48550/arXiv.1908.10084
 229. Rathje S, Mirea DM, Sucholutsky I, Marjeh R, Robertson CE, Van Bavel JJ. GPT is an effective tool for multilingual psychological text analysis. Proceedings of the National Academy of Sciences. 2024 Aug 20;121(34):e2308950121. doi:10.1073/pnas.2308950121

230. De Paoli S. Performing an Inductive Thematic Analysis of Semi-Structured Interviews With a Large Language Model: An Exploration and Provocation on the Limits of the Approach. *Social Science Computer Review*. 2024 Aug 1;42(4):997–1019. doi:10.1177/08944393231220483
231. Bijker R, Merkouris SS, Dowling NA, Rodda SN. ChatGPT for Automated Qualitative Research: Content Analysis. *Journal of Medical Internet Research*. 2024 Jul 25;26(1):e59050. doi:10.2196/59050
232. Xiao Z, Yuan X, Liao QV, Abdelghani R, Oudeyer PY. Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. In: *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces [Internet]*. New York, NY, USA: Association for Computing Machinery; 2023 [cited 2025 Feb 18]. p. 75–8. (IUI '23 Companion). Available from: <https://dl.acm.org/doi/10.1145/3581754.3584136> doi:10.1145/3581754.3584136
233. Chew R, Bollenbacher J, Wenger M, Speer J, Kim A. LLM-Assisted Content Analysis: Using Large Language Models to Support Deductive Coding [Internet]. *arXiv*; 2023 [cited 2025 Feb 19]. Available from: <http://arxiv.org/abs/2306.14924> doi:10.48550/arXiv.2306.14924
234. White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, et al. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT [Internet]. *arXiv*; 2023 [cited 2025 Feb 23]. Available from: <http://arxiv.org/abs/2302.11382> doi:10.48550/arXiv.2302.11382
235. Arlinghaus CS, Wulff C, Maier GW. Inductive Coding with ChatGPT - An Evaluation of Different GPT Models Clustering Qualitative Data into Categories [Internet]. *OSF*; 2024 [cited 2025 Feb 19]. Available from: https://osf.io/gpnye_v1 doi:10.31219/osf.io/gpnye
236. Wang J, Wang C, Luo F, Tan C, Qiu M, Yang F, et al. Towards Unified Prompt Tuning for Few-shot Text Classification [Internet]. *arXiv*; 2022 [cited 2025 Feb 24]. Available from: <http://arxiv.org/abs/2205.05313> doi:10.48550/arXiv.2205.05313
237. Newsroom | TikTok [Internet]. 2019 [cited 2026 Feb 20]. TikTok Joins Internet Matters as Part of its Commitment to Online Safety. Available from: <https://newsroom.tiktok.com/tiktok-joins-internet-matters?lang=en-GB>
238. isolomons. Community Standards Enforcement Report, August 2020. *Meta Newsroom [Internet]*. 2020 Aug 11 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2020/08/community-standards-enforcement-report-aug-2020/>
239. Snap Values [Internet]. 2023 [cited 2026 Feb 21]. Demonstrating respect and empathy on World Kindness Day. Available from: <https://values.snap.com/news/demonstrating-respect-and-empathy-on-world-kindness-day-2023>
240. Vidya Srinivasan. How We Approach Safety and Privacy in Community Chats. *Meta Newsroom [Internet]*. 2022 Sep 13 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2022/09/safety-and-privacy-in-community-chats/>

241. Justin Erlich. Newsroom | TikTok [Internet]. 2023 [cited 2026 Feb 21]. TikTok's state-affiliated media policy. Available from: <https://newsroom.tiktok.com/tiktoks-state-affiliated-media-policy?lang=en>
242. Tara Wadhwa. Newsroom | TikTok [Internet]. 2021 [cited 2026 Feb 21]. New resources to support our community's well-being. Available from: <https://newsroom.tiktok.com/new-resources-to-support-well-being?lang=en>
243. Newsroom | TikTok [Internet]. [cited 2025 Nov 4]. Supporting our Community this #MentalHealthAwareness Month and Year Round. Available from: <https://newsroom.tiktok.com/mentalhealthawarenessmonth-2023?lang=en>
244. Ebony Tucker. Newsroom | TikTok [Internet]. 2021 [cited 2026 Feb 21]. Promoting awareness and safety for sexual assault survivors on TikTok. Available from: <https://newsroom.tiktok.com/promoting-awareness-and-safety-for-sexual-assault-survivors-on-tiktok-ca?lang=en-CA>
245. How We're Supporting People Affected by Eating Disorders and Negative Body Image. Meta Newsroom [Internet]. 2021 Feb 22 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2021/02/supporting-people-affected-by-eating-disorders-and-negative-body-image/>
246. Theo Bertram. Newsroom | TikTok [Internet]. 2020 [cited 2026 Feb 21]. Apoyando a nuestra comunidad durante el Covid-19. Available from: <https://newsroom.tiktok.com/supporting-our-community-through-covid-19-sp?lang=es>
247. Balkissa Idé Sidlo. How Meta Is Preparing for the 2024 South African Elections. Meta Newsroom [Internet]. 2024 Apr 12 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2024/04/how-meta-is-preparing-for-the-2024-south-african-elections/>
248. Fortune Mgwili-Sibanda. Newsroom | TikTok [Internet]. 2024 [cited 2026 Feb 21]. TikTok Reinforces Commitment to Platform Integrity with Key Initiatives Ahead of Ghana Elections. Available from: <https://newsroom.tiktok.com/ghana-elections?lang=en-ZA>
249. bdarwell. Working to Safeguard Elections in Thailand. Meta Newsroom [Internet]. 2019 Jan 30 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2019/01/working-to-safeguard-elections-in-thailand/>
250. Snap Values [Internet]. 2024 [cited 2026 Feb 21]. Planning for the 2024 Elections. Available from: <https://values.snap.com/news/planning-for-the-2024-elections>
251. A Further Update on New Zealand Terrorist Attack. Meta Newsroom [Internet]. 2019 Mar 21 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2019/03/technical-update-on-new-zealand/>

252. Newsroom | TikTok [Internet]. 2023 [cited 2026 Feb 21]. Ensuring a safe, authentic space during times of conflict. Available from: <https://newsroom.tiktok.com/ensuring-a-safe-authentic-space-during-times-of-conflict?lang=en>
253. Helena Lersch. Newsroom | TikTok [Internet]. 2021 [cited 2026 Feb 21]. TikTok's H2 2020 Transparency Report. Available from: <https://newsroom.tiktok.com/tiktok-h-2-2020-transparency-report-sg?lang=en-SG>
254. Johanna Wright. blog.youtube [Internet]. 2017 [cited 2026 Feb 21]. An update on Restricted Mode. Available from: <https://blog.youtube/news-and-events/an-update-on-restricted-mode/>
255. Brie Pegum. Newsroom | TikTok [Internet]. 2024 [cited 2026 Feb 21]. How TikTok counters deceptive behaviour. Available from: <https://newsroom.tiktok.com/how-tiktok-counters-deceptive-behaviour?lang=en-150>
256. Yoel Roth, Ashita Achuthan. X.blog [Internet]. 2020 [cited 2026 Feb 21]. Building rules in public: Our approach to synthetic & manipulated media. Available from: https://blog.x.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media
257. Monika Bickert. Our Approach to Labeling AI-Generated Content and Manipulated Media. Meta Newsroom [Internet]. 2024 Apr 5 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2024/04/metas-approach-to-labeling-ai-generated-content-and-manipulated-media/>
258. Newsroom | TikTok [Internet]. 2024 [cited 2026 Feb 21]. TikTok lanserer dedikert STEM-strøm i Norge og Europa. Available from: <https://newsroom.tiktok.com/stemfeed-norge?lang=nb>
259. Newsroom | TikTok [Internet]. 2022 [cited 2026 Feb 21]. Aportando más contexto a los contenidos en TikTok. Available from: <https://newsroom.tiktok.com/aportando-mas-contexto-a-los-contenidos-en-tiktok?lang=es>
260. Snap Values [Internet]. 2025 [cited 2026 Feb 21]. Enhancing the Safety of our Community by Supporting the TAKE IT DOWN Act. Available from: <https://values.snap.com/news/take-it-down-act>
261. Julie de Baillencourt. Newsroom | TikTok [Internet]. 2023 [cited 2026 Feb 21]. Supporting creators with an updated account enforcement system. Available from: <https://newsroom.tiktok.com/supporting-creators-with-updated-account-enforcement-system?lang=en-150>
262. Tracy Elizabeth. Newsroom | TikTok [Internet]. 2020 [cited 2026 Feb 21]. Supporting youth and families on TikTok. Available from: <https://newsroom.tiktok.com/supporting-youth-and-families-on-tiktok?lang=en>

263. James Beser. blog.youtube [Internet]. 2021 [cited 2026 Feb 21]. A new choice for parents of tweens and teens on YouTube. Available from: <https://blog.youtube/news-and-events/supervised-experiences-for-families-on-youtube/>
264. Adam Mosseri. Introducing Family Center and Parental Supervision Tools on Instagram and in VR [Internet]. 2022 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2022/03/parental-supervision-tools-instagram-vr/>
265. Sandeep Grover. Newsroom | TikTok [Internet]. 2023 [cited 2026 Feb 21]. Introducing a way to refresh your For You feed on TikTok. Available from: <https://newsroom.tiktok.com/introducing-a-way-to-refresh-your-for-you-feed-on-tiktok-us?lang=en>
266. Newsroom | TikTok [Internet]. 2025 [cited 2026 Feb 21]. TikTok utvider fond og tiltak for mental helse og velvære. Available from: <https://newsroom.tiktok.com/no-mental-health-and-mediation-feature?lang=nb>
267. Inspiring Young People in Europe to Make the Internet the Kind of Place They Want to Be. Meta Newsroom [Internet]. 2022 Jun 10 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2022/06/inspiring-young-people-in-europe-to-make-the-internet-the-kind-of-place-they-want-to-be/>
268. James Beser. blog.youtube [Internet]. 2023 [cited 2026 Feb 21]. Continued support for teen wellbeing and mental health on YouTube. Available from: <https://blog.youtube/inside-youtube/continued-support-for-teen-wellbeing-and-mental-health-on-youtube/>
269. Snap Values [Internet]. 2023 [cited 2026 Feb 21]. New features and transparency measures for Snapchatters in the European Union to comply with the Digital Services Act. Available from: <https://newsroom.snap.com/digital-services-act-snap>
270. Antigone Davis. How Meta Addresses Bullying and Harassment. Meta Newsroom [Internet]. 2021 Nov 9 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2021/11/how-meta-addresses-bullying-harassment/>
271. Newsroom | TikTok [Internet]. 2025 [cited 2026 Feb 21]. Stay #safertogether this Safer Internet Day with tips for a positive online experience. Available from: <https://newsroom.tiktok.com/stay-safertogether-this-safer-internet-day-with-tips-for-a-positive-online-experience?lang=en-GB>
272. WhatsApp Channels: Here's Everything You Need To Know. Meta Newsroom [Internet]. 2023 Sep 13 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2023/09/whatsapp-channels-heres-everything-you-need-to-know/>
273. Schroepfer M. An Update on Our Plans to Restrict Data Access on Facebook. Meta Newsroom [Internet]. 2018 Apr 4 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2018/04/restricting-data-access/>

274. Newsroom | TikTok [Internet]. 2025 [cited 2026 Feb 21]. Introducing TikTok Security Checkup. Available from: <https://newsroom.tiktok.com/introducing-tiktok-security-checkup?lang=en>
275. Pavni Diwanji. How Facebook Knows an App User Is Old Enough. Meta Newsroom [Internet]. 2021 Jul 27 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2021/07/age-verification/>
276. Alexandra Evans. Newsroom | TikTok [Internet]. 2020 [cited 2026 Feb 21]. TikTok joins the Alliance to Better Protect Minors Online. Available from: <https://newsroom.tiktok.com/tiktok-joins-the-alliance-to-better-protect-minors-online?lang=en-GB>
277. Snap Values [Internet]. 2023 [cited 2026 Feb 21]. SPS 2023: What's Next for My AI. Available from: <https://newsroom.snap.com/sps-2023-whats-next-for-my-ai>
278. Antigone Davis. Our Approach to Safer Private Messaging. Meta Newsroom [Internet]. 2021 Dec 1 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2021/12/metasp-approach-to-safer-private-messaging/>
279. WhatsApp Channels Are Going Global. Meta Newsroom [Internet]. 2023 Sep 13 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2023/09/whatsapp-channels-global-launch/>
280. Newsroom | TikTok [Internet]. 2025 [cited 2026 Feb 21]. Project Clover update: Enhanced data security with Norwegian data centre fully online. Available from: <https://newsroom.tiktok.com/uk-enhanced-data-security-with-norwegian-data-centre-fully-online?lang=en-GB>
281. blog.youtube [Internet]. 2011 [cited 2026 Feb 21]. Maintaining YouTube Account Security. Available from: <https://blog.youtube/news-and-events/maintaining-youtube-account-security/>
282. New Tools to Help Protect Against Sextortion and Intimate Image Abuse. Meta Newsroom [Internet]. 2024 Apr 11 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2024/04/new-tools-to-help-protect-against-sex-tortion-and-intimate-image-abuse/>
283. Anita Joseph. Verifying the Identity of People Behind High-Reach Profiles. Meta Newsroom [Internet]. 2020 May 28 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2020/05/id-verification-high-reach-profiles/>
284. bdarwell. How Does Facebook Measure Fake Accounts? Meta Newsroom [Internet]. 2019 May 23 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2019/05/fake-accounts/>
285. Del Harvey. X.blog [Internet]. 2018 [cited 2026 Feb 21]. Serving the Public Conversation During Breaking Events. Available from:

https://blog.x.com/en_us/topics/company/2018/Serving-the-Public-Conversation-During-Breaking-Events

286. Adam Presser. Newsroom | TikTok [Internet]. 2024 [cited 2026 Feb 21]. More updates to help the TikTok community create and share safely. Available from: <https://newsroom.tiktok.com/more-updates-to-help-the-tiktok-community-create-and-share-safely?lang=en>
287. Cristos Goodrow. blog.youtube [Internet]. 2021 [cited 2026 Feb 21]. On YouTube's recommendation system. Available from: <https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/>
288. Meta [Internet]. 2023 [cited 2026 Feb 21]. Our approach to explaining ranking | Transparency Center. Available from: <https://transparency.meta.com/features/explaining-ranking>
289. Meta [Internet]. 2025 [cited 2026 Feb 21]. Our Approach to Facebook Feed Ranking| Transparency Center. Available from: <https://transparency.meta.com/features/ranking-and-content/>
290. Nick Clegg. How AI Influences What You See on Facebook and Instagram. Meta Newsroom [Internet]. 2023 Jun 29 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2023/06/how-ai-ranks-content-on-facebook-and-instagram/>
291. Adam Mosseri. About.Instagram [Internet]. 2023 [cited 2026 Feb 21]. Instagram Ranking Explained | Instagram Blog. Available from: <https://about.instagram.com/blog/announcements/instagram-ranking-explained>
292. Facebook help center [Internet]. [cited 2026 Feb 21]. Customize your Facebook Feed by selecting Interested or Not interested on a post | Facebook Help Center. Available from: <https://www.facebook.com/help/231515559920930>
293. Newsroom | TikTok [Internet]. 2021 [cited 2026 Feb 21]. An update on our work to safeguard and diversify recommendations. Available from: <https://newsroom.tiktok.com/an-update-on-our-work-to-safeguard-and-diversify-recommendations?lang=en>
294. Newsroom | TikTok [Internet]. 2023 [cited 2026 Feb 21]. Una actualización sobre el cumplimiento de nuestro compromiso en relación a la Ley de Servicios Digitales. Available from: <https://newsroom.tiktok.com/actualizacin-sobre-el-cumplimiento-de-nuestro-compromiso-en-relacion-a-la-ley-de-servicios-digitales?lang=es>
295. Snap Newsroom [Internet]. 2020 [cited 2026 Feb 21]. Rewarding Creativity on Spotlight: Shining a Light on the Best Snaps. Available from: <https://newsroom.snap.com/rewarding-creativity>

296. Susan Wojcicki. blog.youtube [Internet]. 2022 [cited 2026 Feb 21]. Letter from Susan: Our 2022 Priorities. Available from: <https://blog.youtube/inside-youtube/letter-susan-our-2022-priorities/>
297. Mark Zuckerberg. Mark Zuckerberg - Continuing our focus for 2018 to make... | Facebook [Internet]. 2018 [cited 2026 Feb 21]. Available from: <https://www.facebook.com/zuck/posts/continuing-our-focus-for-2018-to-make-sure-the-time-we-all-spend-on-facebook-is-/10104445245963251/>
298. Giving People More Control on Instagram and Facebook. Meta Newsroom [Internet]. 2021 May 26 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2021/05/giving-people-more-control/>
299. Continuing to Create Age-Appropriate Ad Experiences for Teens. Meta Newsroom [Internet]. 2023 Jan 10 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2023/01/age-appropriate-ads-for-teens/>
300. Newsroom | TikTok [Internet]. 2022 [cited 2026 Feb 21]. Furthering our advertising transparency efforts: introducing “About this ad.” Available from: <https://newsroom.tiktok.com/furthering-our-advertising-transparency-efforts-introducing-about-this-ad-ca?lang=en-CA>
301. Snap Values [Internet]. 2023 [cited 2026 Feb 21]. Updating our Advertising in Europe and the UK. Available from: <https://values.snap.com/news/updating-our-advertising-in-europe-and-uk-2023>
302. Newsroom | TikTok [Internet]. 2024 [cited 2026 Feb 21]. Fostering mental well-being awareness and literacy with the World Health Organization. Available from: <https://newsroom.tiktok.com/fostering-mental-wellbeing-awareness-and-literacy-with-the-world-health-organization?lang=en>
303. Susan Wojcicki. blog.youtube [Internet]. 2021 [cited 2026 Feb 21]. Letter from Susan: Our 2021 Priorities. Available from: <https://blog.youtube/inside-youtube/letter-from-susan-our-2021-priorities/>
304. Eric Han. Newsroom | TikTok [Internet]. 2020 [cited 2026 Feb 21]. Countering hate on TikTok. Available from: <https://newsroom.tiktok.com/countering-hate-on-tiktok?lang=en>
305. Nick Clegg. Meta Launches New Content Moderation Tool as It Takes Chair of Counter-Terrorism NGO. Meta Newsroom [Internet]. 2022 Dec 13 [cited 2026 Feb 21]. Available from: <https://about.fb.com/news/2022/12/meta-launches-new-content-moderation-tool/>
306. Newsroom | TikTok [Internet]. 2025 [cited 2026 Feb 21]. TikTok anima a su comunidad mantenerse segura en el Día de Internet Segura con consejos para una experiencia en línea positiva. Available from: <https://newsroom.tiktok.com/tiktok-dia-de-internet-seguro-2025?lang=es>

307. Snap Values [Internet]. 2024 [cited 2026 Feb 21]. New Features to Help Protect Our Community. Available from: <https://values.snap.com/news/new-features-to-help-protect-our-community>
308. Meta AI [Internet]. 2023 [cited 2026 Feb 21]. The AI behind unconnected content recommendations on Facebook and Instagram. Available from: <https://ai.meta.com/blog/ai-unconnected-content-recommendations-facebook-instagram/>
309. Transparency Center [Internet]. [cited 2026 Feb 14]. Available from: <https://transparency.meta.com/reports/>
310. Edelson L, Kovba B, Yershova H, Botelho A, McCoy D, Lauinger T. Measurement and Metrics for Content Moderation: The Multi-Dimensional Dynamics of Engagement and Content Removal on Facebook. *Journal of Online Trust and Safety*. 2025 Apr 28;2(5). doi:10.54501/jots.v2i5.220
311. Constine J. Facebook's U.S. user count declines as it prioritizes well-being. *TechCrunch* [Internet]. 2018 Jan 31 [cited 2025 Nov 30]. Available from: <https://techcrunch.com/2018/01/31/facebook-time-spent/>
312. Guess AM, Malhotra N, Pan J, Barberá P, Allcott H, Brown T, et al. How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*. 2023 Jul 28;381(6656):398–404. doi:10.1126/science.abp9364
313. MIT Media Lab [Internet]. [cited 2026 Feb 14]. Introducing MIT Media Lab's Advancing Humans with AI (AHA) Research Program. Available from: <https://www.media.mit.edu/posts/introducing-aha/>
314. Ivey R, Sullivan E, Lomas T, Lee M, Pawelski J, Sanders M, et al. Human Flourishing: A Guiding Principle for AI Systems Design? [SSRN Scholarly Paper] [Internet]. Rochester, NY: Social Science Research Network; 2024 [cited 2026 Feb 14]. Available from: <https://papers.ssrn.com/abstract=5185045> doi:10.2139/ssrn.5185045
315. VanderWeele TJ, Teubner JD. Flourishing Considerations for AI. *Information*. 2026 Jan;17(1):88. doi:10.3390/info17010088
316. Cohen J. A power primer. *Psychological Bulletin*. 1992;112(1):155–9. doi:10.1037/0033-2909.112.1.155
317. Gautam S, Chandra M, De A, Chakravorti T, Malik G, De Choudhury M. arXiv.org [Internet]. 2025 [cited 2026 Feb 20]. Towards Experience-Centered AI: A Framework for Integrating Lived Experience in Design and Development. Available from: <https://arxiv.org/abs/2508.06849v1>
318. Fang CM, Liu AR, Danry V, Lee E, Chan SWT, Pataranutaporn P, et al. How AI and Human Behaviors Shape Psychosocial Effects of Extended Chatbot Use: A Longitudinal Randomized Controlled Study [Internet]. arXiv; 2025 [cited 2026 Feb 20]. Available from: <http://arxiv.org/abs/2503.17473> doi:10.48550/arXiv.2503.17473

319. Zimmerman A, Janhonen J, Beer E. Human/AI relationships: challenges, downsides, and impacts on human/human relationships. *AI Ethics*. 2024 Nov 1;4(4):1555–67. doi:10.1007/s43681-023-00348-8
320. Ghai S, Magis-Weinberg L, Stoilova M, Livingstone S, Orben A. Social media and adolescent well-being in the Global South. *Current Opinion in Psychology*. 2022 Aug 1;46:101318. doi:10.1016/j.copsyc.2022.101318
321. Viswanath K, McCloud RF, Lee EWJ, Bekalu MA. Measuring What Matters: Data Absenteeism, Science Communication, and the Perpetuation of Inequities. *The ANNALS of the American Academy of Political and Social Science*. 2022 Mar 1;700(1):208–19. doi:10.1177/00027162221093268