Understanding unreflective habit formation on TikTok as a system safety problem: an explorative study in Vietnam

Technology, Policy and Management

Ms Engineering and Policy Analysis

Author: Nhi Tran
Student Number: 5914884
Date of Defense: 2/7/2025

ŤUDelft

Thesis Report

Understanding unreflective habit formation on TikTok as a system safety problem: an explorative study in Vietnam

Submitted by

Nhi Tran

5914884

Graduation:

July 2025

Supervisors:

Dr.ir. Roel I.J. Dobbe

Dr.ir. Lavinia Marin

Chair:

Dr. Haiko G. van der Voort

A thesis report submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of Master of Science in Engineering and Policy Analysis

The work in this thesis was carried out in the:

Engineering and Policy Analysis
Delft University of Technology

Executive Summary

There are increasing concerns regarding a more insidious consequence of social media: the erosion of user autonomy through manipulative design strategies (Susser, Roessler, and Nissenbaum 2019; Norlock 2021; Sahebi and Formosa 2022). Social media platforms not only encourage consumerism by subtly influencing users to make purchases without full awareness (Mik 2016), but also contribute to political manipulation and opinion shaping, thereby undermining democratic processes (Susser, Roessler, and Nissenbaum 2019; Sahebi and Formosa 2022). Thus, the loss of user autonomy presents a threat not only at the individual level but also at the societal level. This research focuses on unreflective habit formation, a process through which user autonomy is undermined. The aim is to explore this phenomenon in the context of TikTok usage in Vietnam and to apply System-Theoretic Accident Model and Processes (STAMP) model to conceptualize the issue and the System-Theoretic Process Analysis (STPA) method to provide actionable recommendations to the Vietnam Government.

The study begins with the following research question:

> **RQ 1:** To what extent does social media enforce unreflective habit formation?

Desk research serves as the primary method to answer this question. By integrating various accounts of autonomy - procedural, weak substantive, strong substantive, and relational - as articulated by Stoljar (2013), with the Human-Computer Interaction (HCI) perspective on autonomy from Bennett et al. (2023), the research identifies five key dimensions of autonomy:

- **Decision**: To what extent do users engage in critical internal reflection before acting in response to stimuli from social media?

- **Execution**: To what extent do users intentionally carry out the action in response to the stimuli?

- **Self-congruence**: To what extent do users' actions and decisions align with their values and goals?

- **Interdependence**: To what extent are users' social relationships constructive or destructive to their autonomy?

- **System hostility**: To what extent do social media systems impede user autonomy?

Social media exploits neurological reward cycles to reinforce user habits (Esposito and Ferreira 2024). Among the five dimensions, three are found to be undermined. First, the decision dimension is violated as user behavior becomes an automatic reaction to stimuli without critical reflection. Second, execution is undermined due to the absence of intentionality. Third, the hyper-stimulating and unpredictable rewards foster a system hostility that obstructs autonomous behaviors. When these conditions lead users to adopt automatic routines, the result is defined as unreflective habit formation.

The research proceeds with:

RQ 2a: To what extent is using TikTok an unreflective habit for Vietnamese female university students?

Semi-structured interviews were conducted with 12 Vietnamese female university students in Hue during April 2025. The interviews reveal that their habit of using TikTok developed outside of conscious awareness. TikTok is perceived as a benevolent companion offering excitement, social connection, knowledge, beauty tips, and the chance for virality. Given these perceived benefits, users rarely reflect critically on TikTok's influence. Instead, their engagement is steered by deeply embedded reward cycles, reinforcing unreflective habit formation.

This leads to the next question:

RQ 2b: What are the resulting losses associated with this habit?

Three major losses are identified. First, users lose valuable time, as TikTok's design impedes their ability to regulate screen time. Its management features are symbolic rather than practical. Second, users lose the capacity for sustained attention, impairing their ability to engage in deep cognitive tasks. This is due to constant partial attention to TikTok stimuli and the brevity of its content, which restructures users' cognitive patterns toward shallow engagement (Norlock 2021; Carr 2011). Third, highly personalized TikTok feeds limit exposure to diverse role models, potentially leading users to internalize their online personas (Attrill-Smith 2019).

The research proceeds with:

RQ 3a: Using STAMP model and STPA method, what are the recommendations to prevent the identified losses to Vietnamese female university students?

The responsibility for regulating social media platforms like TikTok lies primarily with the Vietnam Government. The Vietnam Government should adopt a combined strategy that integrates strict regulation of social media service providers with comprehensive public education. These measures must be informed by up-to-date scientific research on user behavior and platform design, ensuring that policies remain relevant and effective. Importantly, regulations should not target TikTok in isolation, as the platform's features are widely replicated across the industry. Instead, all social media platforms operating in Vietnam must be held equally accountable to promote uniform safety standards and prevent harmful design practices from being shifted or reproduced elsewhere.

Lastly, the research addresses:

RQ 3b: To what extent are STAMP model and STPA method suitable for this problem?

In summary, applying the STAMP model and STPA method to social media reveals both challenges and benefits. Conceptualizing pernicious harms proved difficult, leading this research to shift focus from losses to the process of unreflective habit formation, emphasizing the prevention of repeated impulsive use. Some causal categories of loss scenarios in STPA were also less relevant for identifying loss scenarios. Despite these challenges, the control structure provides a systematic overview, while Unsafe Control Actions (UCAs), controller constraints, and loss scenarios offer detailed, actionable insights that support targeted and comprehensive policy strategies.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my family. Their unwavering emotional support and encouragement have carried me through the most challenging moments of this journey. Thanks to their belief in me and their readiness to always have my back, I was able to complete this thesis in a remarkably short amount of time.

I am sincerely thankful to my supervisors for their insightful guidance and continuous support throughout this research. Their recommendations—through paper suggestions and critical discussions—not only helped me gain clarity on my research direction but also enabled me to identify and overcome key obstacles that had previously kept me at a standstill.

I would also like to thank my chair for their constant positivity and encouragement. Their ability to connect my work to the broader goals of my study program helped ground my research within the academic framework, offering both structure and purpose.

Finally, I am grateful to my friends who have shown me kindness and encouragement. Their warmth and steady reassurance have been a source of strength and motivation along the way.

Contents

List of Figures

List of Tables

Acronyms

1. Introduction

1.1. Grand Challenge

Social media platforms are inherently attractive to users across all ages, genders, and races (Kaplan and Haenlein 2010). As of February 2025, there are approximately 5.24 billion social media users worldwide - equivalent to 63.9% of the global population (Statista 2025b). With 5.56 billion internet users in the same period (Statista 2025b), it is fair to say that nearly every internet user also uses social media. The five most popular platforms - Facebook, YouTube, Instagram, WhatsApp, and TikTok - each having between 1 and 3 billion monthly active users (Statista 2025a).

With such broad reach, social media brings both benefits and risks. For example, Kubheka, Carter, and Mwaura (2020) found that social media can effectively and affordably promote health programs in South Africa by rapidly spreading information and enabling virtual engagement despite geographic barriers. However, misinformation and uneven levels of health and digital literacy pose significant challenges to these applications (Kubheka, Carter, and Mwaura 2020). For another instance, while social media has served as a tool for feminist advocacy, misogynistic content remains more prevalent, in part due to algorithmic amplification (Boyle and Berridge 2023). It is no longer surprising that social media platforms can detect and respond to users' emotions in real time based on the vast datasets they collect (Levin 2017). While artificial intelligence has shown some promise in diagnosing mental health issues through user data, its effectiveness remains contested (Laacke et al. 2021). In contrast, using AI to predict consumer behavior is already a common practice (Mik 2016; Levin 2017; Gal 2018; Anker 2024).

Beyond these applications and challenges, scholars increasingly point to a more insidious effect of social media: the erosion of user autonomy (Norlock 2021; Sahebi and Formosa 2022). Susser, Roessler, and Nissenbaum (2019) argues that technologies designed to manipulate users can directly undermine their autonomy. Sahebi and Formosa (2022) offer a detailed explanation of how this occurs. First, by exploiting users' desire for social contact and fear of missing out, platforms incentivize them to surrender their data, which is then used to build detailed digital profiles. Second, these profiles feed personalized content back to users, subtly shaping their thoughts and behaviors beyond conscious awareness (Susser, Roessler, and Nissenbaum 2019). Third, interface design elements manipulate user behavior, encouraging continuous engagement.

The Cambridge Analytica scandal exemplifies how extensive user data can be used to manipulate political opinions, with significant consequences for democratic processes (Susser, Roessler, and Nissenbaum 2019). Similar concerns have arisen over Twitter's (now X) algorithm amplifying particular political groups (Sahebi and Formosa 2022), and over Facebook's ability to influence users' emotional states in ways that benefit advertisers (Susser, Roessler, and Nissenbaum 2019). These cases illustrate how technology's persuasive capacity, when combined with exploitative intent, leads to manipulation and loss of autonomy.

Despite such threats, user reactions vary. Some try to subvert algorithms by intentionally interacting with content they do not actually enjoy, tricking the system into adjusting recommendations (Karizat et al. 2021). Others remain unaware or indifferent. Even in highly digitalized societies, many users do not understand how algorithmic systems work (Sahebi and Formosa 2022). As a result, they may unknowingly submit to manipulation (Malik et al. 2022), or believing that personalized content is an indication that the algorithms care about them and their preferences (Savolainen and Ruckenstein 2024).

Still, some scholars challenge the view that autonomy is entirely lost. Klenk and Hancock (2019) argue that autonomy is a complex concept, and critiques like that of Susser, Roessler, and Nissenbaum (2019) rely heavily on externalist definitions. Given the philosophical diversity in how autonomy is defined (Stoljar 2013), Klenk and Hancock (2019)'s critique may be valid. Nonetheless, if one adopts an externalist framework, the empirical examples cited by Susser, Roessler, and Nissenbaum (2019) do suggest significant threats to autonomy.

Other scholars argue that algorithmic assistance may enhance autonomy. For instance, delegating minor decisions to algorithms can free mental resources for more important ones (Gal 2018). However, even seemingly user-aligned recommendations are based on algorithmic approximations of preference, raising concerns about the transparency and control of decision-making processes (Gal 2018). Imagine a user looking for a café to meet friends. Rather than sorting through countless posts, they are shown a manageable shortlist, easing cognitive load. While this is helpful, two critical issues emerge. First, users rarely choose to configure these algorithms. Even in advanced societies like Germany and Norway, most people lack awareness of how such systems operate (Savolainen and Ruckenstein 2024). Users are nudged toward outcomes that benefit platforms, often without knowing it (Malik et al. 2022). Second, limiting users' exposure to alternative perspectives constrains their autonomy. When algorithms curate content, users may miss opportunities to encounter diverse worldviews or form independent values (Malik et al. 2022; Anker 2024). While trivial decisions like choosing a café may have limited implications, choices about news consumption or political beliefs can carry much greater weight. As the Cambridge Analytica case demonstrates, such influence can distort voting behavior - a deeply consequential expression of personal autonomy.

1.1.1. unreflective habit formation

The loss of user autonomy is a vast field. To make this study tractable, the research focuses on one specific manifestation: unreflective habit formation. This refers to the development of habits without conscious awareness or intentional acceptance (Marin 2025). While individuals may maintain distinct online and offline personas, the former can eventually shape the latter - especially when performed repeatedly - due to the malleability of personal identity (Attrill-Smith 2019). Over time, these habits may reshape one's identity, leading to identity rigidification (Marin 2025). Although transformation can be positive, social media can also drive undesirable changes without users' knowledge or consent (Norlock 2021). In such cases, individuals lose the opportunity to reflect on or influence who they become.

People naturally evolve, often through books, friendships, or cultural exchange. So why is unreflective habit formation a concern? The answer lies in the scale and frequency of social media use. The global average is 18 hours and 41 minutes per week - over 2.5 hours per day (DataReportal 2025). Time spent on social media could otherwise be directed toward meaningful or impactful activities. Moreover, platforms exploit user attention for commercial gain

(Mik 2016; Susser, Roessler, and Nissenbaum 2019; Sahebi and Formosa 2022), increasing the cost for those who develop compulsive usage or purchasing habits. Unreflective habit formation also has broader societal consequences. It has been linked to reduced attention spans (Norlock 2021), emotional manipulation for advertising, and political indoctrination (Susser, Roessler, and Nissenbaum 2019; Sahebi and Formosa 2022; Malik et al. 2022) - each of which threatens individual well-being and social cohesion.

Despite the severity of these effects, banning social media is not a viable solution. This is because these platforms have undeniable benefits, such as enabling community-based health interventions. On the other hand, expecting users to self-regulate is unrealistic. There are three main reasons for this. First, users are unaware that these habits are forming. By definition, unreflective habit formation occurs beneath conscious awareness. Even though there are settings seemingly designed to help users control their usage, they are often disabled by default - and especially inaccessible to users with lower digital competence (Mik 2016). Second, even social media engineers have admitted difficulty in resisting the very platforms they helped build (Lewis 2017). Some must resort to external controls, such as delegation to others (Lewis 2017). This is because social media is designed with "ludic loops" that reinforce compulsive engagement (Norlock 2021). Third, leaving social media is socially costly. These platforms satisfy deep-seated needs for social contact, forcing users to choose between connection and disconnection (Sahebi and Formosa 2022). Avoiding such alluring charms promised by social media is challenging (Norlock 2021).

In summary, unreflective habit formation is an alarming expression of diminished user autonomy. It harms both individuals and society. Since social media provides value to many, banning it negatively affect those benefiting from it. Yet, expecting users to consciously avoid unreflective habits is implausible.

1.2. Policy Problem and Project Scope

Regulating social media to mitigate the loss of user autonomy presents significant challenges. One key issue is the tendency to apply outdated policy solutions to emerging technologies, despite the fact that such legacy frameworks were developed to address fundamentally different problems (Obar and Wildman 2015). Moreover, policymakers often lack sufficient technological literacy, which impairs their ability to develop effective regulations (Obar and Wildman 2015; Mik 2016).

Another major challenge lies in the flawed assumptions underlying existing legal frameworks. Many digital policies presuppose that individuals are fully competent to safeguard their own interests (Susser, Roessler, and Nissenbaum 2019). This includes the belief that providing users with more information will enable them to make better-informed decisions, thereby enhancing their autonomy (Mik 2016). In practice, however, users are already inundated with information in the digital environment, which significantly diminishes their capacity to absorb and act upon disclosures - particularly when it concerns manipulative platform practices (Mik 2016).

Another challenge is that there is not enough scrutiny over the social media business model, which profit from advertisement revenues through holding users' attention captive by giving rewards using recommendation algorithms and engagement-boosting features (Marcus and Koester 2024). It is precisely this business model that drives social media providers to relentlessly maximize user engagement at all costs (Lewis 2017; Sahebi and Formosa 2022).

Since the business model pushes for the evolvement of user engagement mechanisms, which substantially infringe on users autonomy, it should be viewed as one of the root causes of the loss of user autonomy. As long as the root causes, which is the threat to users autonomy, is not addressed, no effective remedies will be found (Susser, Roessler, and Nissenbaum 2019).

This research seeks to assist policymakers in conceptualizing and addressing the problem of unreflective habit formation, a specific dimension of the broader issue of loss of user autonomy. Vietnam is selected as the research focus due to its large and active social media user base and the relative scarcity of academic studies addressing social media's impact in the country (Nguyen et al. 2025). Additionally, the Vietnamese government has demonstrated a strong interest in regulating social media platforms to promote economic integration, while simultaneously maintaining strict control over platform operations - even those run by foreign, Western-based companies (Le and Hutchinson 2022). Importantly, Vietnamese authorities have shown their strong intentions to compel social media providers to comply with national directives (Le and Hutchinson 2022). Given the intention of the Vietnam Government in managing social media, the research hopes to assist this effort with insights for viable recommendations.

TikTok is a social media platform centered on the creation and dissemination of short-form videos (TikTok 2025a). Despite being a new comer to the scene, it is rapidly rising in popularity among Generation Z (Gen Z) users in Vietnam, surpassing incumbent players like Facebook and Instagram (Huynh 2025). Given the popularity of TikTok in Vietnam, this research then focuses on this social media platform.

Members of Gen Z are defined as those born between 1995 and 2012 (aged 13 to 28). Segmenting Gen Z by educational level, there are four main groups: secondary school (ages 13–16), high school (16–18), university students (18–21), and university graduates (21+). Although family and social contexts differ among each individual in these groups, educational stage and age already introduce significant variance. To ensure internal consistency and meaningful analysis, this study will narrow its scope to university students aged 18–21, based on three main reasons. Firstly, considering children vulnerable to harms from TikTok, the Vietnam Government strictly requires TikTok to take measures to protect this group of users (Thu Hang 2023). Since the Government already takes close watch on this group of users, the research will look into the age group 18 and older. Secondly, since age 18 is commonly considered the threshold of psychological maturity (Icenogle et al. 2019), users in this age group can be considered capable of exercising their autonomy through cognitive processes. Thirdly, university students share a relatively homogeneous environment in terms of lifestyle and educational exposure, unlike graduates whose experiences become more diverse and less controllable for research purposes. Furthermore, given that social media usage patterns differ between male and female users (Park et al. 2023), the research focuses on female students to reduce variability resulting from this difference.

In Vietnam, the two most influential actors in shaping the social media environment are the platform providers and the government (Le and Hutchinson 2022). However, both primarily concentrate on content and account regulation, rather than on mitigating unreflective habit formation. For instance, TikTok's public safety documentation emphasizes content moderation - removing harmful materials - but makes little mention of how platform design may encourage habitual use (TikTok 2024; Keenan 2021). Similarly, in the 2023 report by Ministry of Information and Communication (MIC), TikTok's regulatory violations in Vietnam are discussed almost exclusively in relation to content-related concerns (Ministry of Science and Technology 2023). Figure 1.1 provides an overview of the actors and their interactions in

the social media socio-technical system in Vietnam pertaining to the topic and scope of the project. Their interactions are furthered explained in other sections of this report.
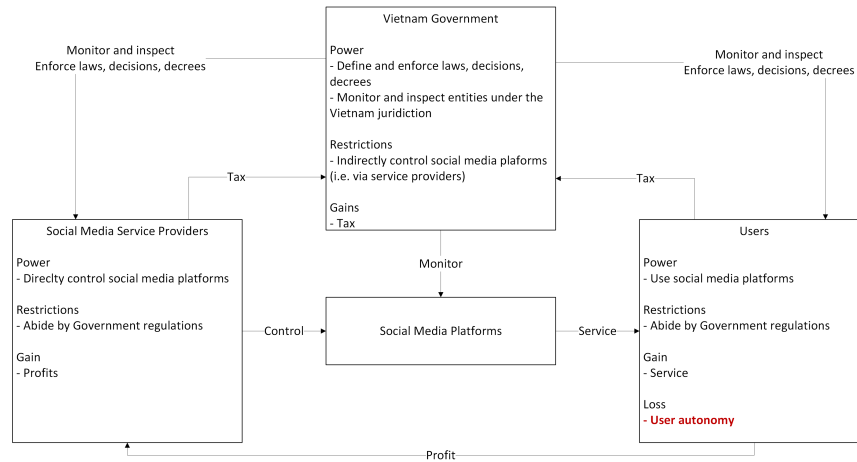
Figure 1.1.: Main actors in social media socio-technical system in Vietnam

The researcher believes that with a more structured definition of the problem, actionable recommendations can be identified. To bring actionable structure to the issue, the researcher aims to apply Systems Engineering frameworks to conceptualize the problem. The research select STAMP as the conceptual model and STPA as the method to demarcate the problem. These two concepts are explained in more details in the next Chapter.

In summary, although social media platforms pose considerable threats to user autonomy, this issue remains underexplored in both academic literature and regulatory discourse - especially in the Vietnamese context. Focusing on unreflective habit formation - a consequence of loss of user autonomy, the research aims to offer actionable insights that may inform future regulatory approaches in Vietnam.

1.3. Research Questions

The research first addresses the following research question:

- **RQ 1: To what extent does social media enforce unreflective habit formation?**

Bringing the discussion into the context of Vietnam for female university students, the research addresses the following research question:

- **RQ 2a: How is using TikTok an unreflective habit for Vietnamese female University students?**

- **RQ 2b: What are the resulting losses associated with this habit?**

The findings from RQ 2, especially the identified losses, are the necessary inputs to the next research question, which is as follows:

- **RQ 3a: Using** STAMP **model and** STPA **method, what are the recommendations to prevent the identified losses to Vietnamese female university students?**

- **RQ 3b: To what extent are** STAMP **model and** STPA **method suitable for this problem?**

The structure of the report is as follows: Section 2 introduces the STAMP model - including key concepts and constructs, and the STPA method - including analytical steps and outputs. Section 3 outlines the research design and data collection strategy. Chapter 4 outlines the theoretical concepts that answer RQ 1. In particular, Section 4.1, 4.2 defines autonomy in the context of HCI. Section 4.3 situates loss of user autonomy in the context of social media. Section 4.4 explains how unreflective habit formation occurs on social media. The interview findings are presented in Chapter 5, which provides evidence to answer RQ 2. Section 6 is the analysis of the results to answer RQ 2. Section 7 applies the STAMP model to conceptualize the problem and the STPA method to generate actionable insights to answer RQ 3. Chapter 8 summarizes the findings from the STPA analysis in the form of policy recommendations to answer RQ 3a. Chapter 9 outlines the challenges and benefits from using the STAMP model and the STPA method on unreflective habit formation on TikTok to answer RQ 3b. The report concludes with suggestions for future improvements.

2. STAMP and STPA

2.1. Systems Theory and System Safety

The core principle of Systems Theory, which provides a theoretical foundation for System Engineering, states that as a system is more than the sum of its individual components, a system must be viewed in its wholeness. In Systems Theory, safety is an emergent property because it arises from component and subsystem interactions. So, to maintain the safety of the system, merely ensuring the safety of system components separately is not enough. Rather, it must be viewed in the context of components interactions in the system in its completeness. Furthermore, since emergent properties can be controlled by applying constraints on individual components, improving safety is a matter of implementing safety controls, either by human or automated controllers, or both (Leveson 2012).

System Safety is a subdiscipline of System Engineering. Following the System Engineering safety approach, there are two main assumptions guiding System Safety. Firstly, individual component behaviors cannot be understood out of context of the system. Secondly, due to the complex and nonlinear interactions among the components and subsystems, system optimization cannot be separately conducted on individual components or subsystem because these siloed attempts might worsen the overall system performance (Leveson 2012).

2.2. STAMP

Systems Theory establishes that system safety is a matter of enforcing controls and safety-constraints of the behaviors and interactions of system components through three areas: physical design, process (operations, maintenance, etc.), and social control (organizational management, governmental regulations, self-interests, etc.). Any mistake in implementing controls or safety constraints in any of these three areas will result in accidents (Leveson 2012).

STAMP, developed based on Systems Theory, is a model that can be used to enhance System Safety. There are two principles guiding STAMP that is built on Systems Theory. Firstly, identifying causes of ineffective control is essential in investigating the causes of accidents. Secondly, safety control and constraints must be included in the physical design, process and social control of a system to prevent future accidents (Leveson 2012).

There are two foundational concepts of STAMP, namely:

- **Accident** — Leveson (2012, p.181) defines an accident as "An undesired or unplanned event that results in a loss, including loss of human life or human injury, property damage, environmental pollution, mission loss, etc."

- **Hazard** — Leveson (2012, p.184) defines a hazard as "A system state or set of conditions that, together with a particular set of worst-case environmental conditions, will lead to an accident (loss)."

Given the definition of accident by Leveson (2012), loss encompasses undesirable consequences that put a subject at a worse state than it used to be before the accident. The subject can be human or non-human, tangible or intangible. Losses resulting from hazards in technical systems often are measurable. However, the harms of unreflective habit formation are rather pernicious in several ways. Firstly, the agent herself is unaware of the subtle and gradual transformation she experiences. Secondly, in case that she eventually re-evaluates her new identity, given the hostile environment she is in, she might yield to the environmental influence and accept her new identity. Since the agent herself is unaware of or subdued by the incremental changes caused by the system, it is hard to assert that she has been harmed. This problem will be further addressed in the section 7.1.1.

There are three main constructs in STAMP: safety constraints, hierarchical safety control structures and process models (Leveson 2012):

- Safety constraints: the set of system boundaries to prevent the system from falling into unsafe states. Safety constraints guide the design and enforcement of safety controls in a system (Leveson 2012).

- Hierarchical safety control structures: represent the hierarchy of the socio-technical system (government, organization, physical system, etc.). Each level of the safety control structure imposes safety control actions and receives feedback from the level immediately beneath it (Leveson 2012).

- Process model: the mental representation of the process being control of each controller, both human and automated (Leveson 2012).

2.3. STPA

STPA is a new hazard analysis technique based on STAMP model. In addition to the physical/mechanical components of the system, STPA considers other causal factors identified in STAMP that are omitted by older hazard analysis techniques, such as design errors, hazardous component interactions, human controller's decision-making errors and social and organizational unsafe behaviors, etc. (Leveson 2012).

As outlined by Leveson and Thomas (2018), there are four steps in an STPA analysis:

1. Define the purpose of the analysis by identifying losses, hazards, and system-level constraints.

2. Model the control structure by deriving a conceptual model of the system showing the controlled processes, controllers, control actions, and feedback information.

3. Identify the UCA. A control action can give rise to hazardous situations in four main ways:

    - Not providing the control action results in hazards

    - Providing the control action results in hazards

- The control action being applied too early, too late, or out of order results in hazards

- The control action being applied too long or stopped too soon results in hazards

4. Identify loss scenarios, which describe the causal factors resulting in UCA and losses. There are four categories of causal factors:

   - Unsafe controller behaviors: Factors pertaining to the controllers, such as physical failure, inadequate control algorithms or process models, and unsafe control input.

   - Inadequate feedback and information: Factors causing feedback information to be improperly received or transmitted.

   - Issues with the controlled paths: Factors related to faulty actuators that lead to improperly executed commands or erroneous feedback to sensors and, thus, controllers.

   - Issues with the controlled processes: Factors related to the malfunctioning of the controlled processes, causing improper receipt or execution of control actions, or inaccurate status responses to controllers.

2.4. Literature Review: Applications of STAMP and STPA in socio-technical systems

The literature review process for this research was conducted systematically to ensure comprehensive and relevant results. Google Scholar and Scopus are selected as the primary repositories due to their vast pool of academic articles. The search phrases for literature review are **("social media" AND "intervention") AND ("systems theory" OR STPA OR STAMP)**. The search period is limited to 2003 and 2025, around the time social media started to gain traction with the launch of Facebook.

There are 2696 articles matching the search phrases on Google Scholar. Using Publish and Perish software (Harzing 2007), the results are exported into CSV files. On Scopus database, these search phrases yield 0 articles, even after adding additional key words such as "safety". The result CSV files are then processed in a Jupyter Notebook (Kluyver et al. 2016) to identify relevant articles by applying filters using the search phrases on the respective columns of interests of these files.

There is 1 article from Google Scholar that contain the search phrases in the columns "Title" and "Abstract". Additionally, a prior review by Patriarca et al. (2022) was referenced, which included a list of 321 papers focused on the applications of STAMP, STPA, and CAST from 2002 to 2021. By filtering this list to include studies within the industries of "Information Technology", "Other" and "Other (Specify)", 8 relevant papers were identified.

In total, the literature review resulted in 8 articles from Patriarca et al. (2022) and 1 article from the original searches, as summarized in Table 2.1. The literature review process is summarized in Figure 2.1. All 9 articles provide a case study in applying STAMP or STPA. However, none of them use social media in their case studies, which is shown in the column "Application" in Table 2.1. The literature review indicates that despite its popularity and applications in a vast industry, STAMP and STPA are yet to be applied to social media. This is

Figure 2.1.: Literature review process
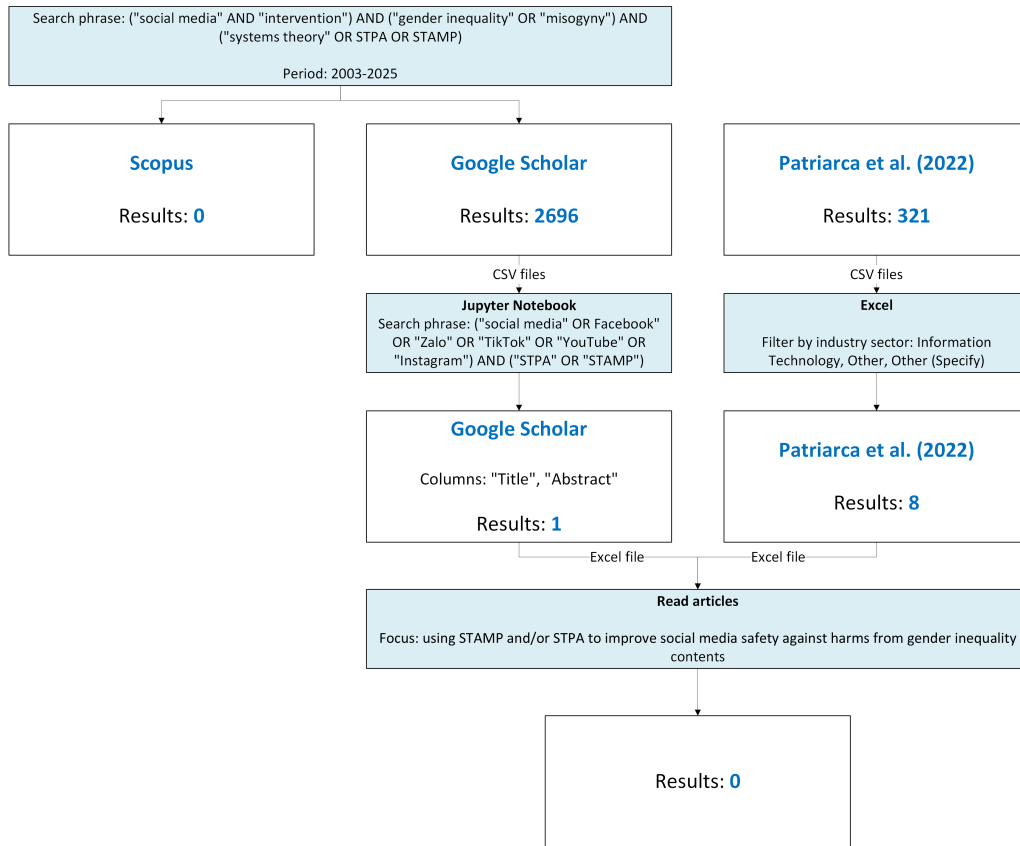
in accordance with the literature review conducted by Zhang et al. (2022) on the application of STAMP and STPA.

2.5. Project Approach

The Systems Theory principles can be applied to ensure safety for not only engineering systems, but also social ones (Leveson 2012). In fact, STAMP and STPA have proved to be pragmatic for socio-technical system safety analysis and design in different industries (Patriarca et al. 2022; Zhang et al. 2022). As social media is perceived as a socio-technical system, STAMP and STPA can be applied to design safe social media environment for users.

There are bases for these claims. The 3 constructs of STAMP model implies a comprehensive approach to understand the problem.

- The set of safety constraints resulting from STAMP can provide materials to design safety measures for social media.

- The process models of different controllers can shed lights on sources of hazards and remedies.

Title	Authors	Application
An exploratory study on applying a scrum development process for safety-critical systems	Wang, Ramadani, and Wagner (2017)	Smart home
Enriching Systems Theory Based Cyber-Security and Safety Analysis Using Stakeholder Value Networks	Sidhu and Moulton (2020)	Mining system
Privacy Risk Analysis Based on System Control Structures: Adapting System-Theoretic Process Analysis for Privacy Engineering	Shapiro (2016)	Smart TV
STAMP-based Approach to Analyze Safety, Security and Data Privacy	De Souza et al. (2019)	Electronic voting system
Towards Combined Safety and Security Constraints Analysis	Pereira et al. (2017)	Revolving door
Using hazard analysis STAMP/STPA in developing model-oriented formal specification toward reliable cloud service	Hata et al. (2015)	Cloud service
STPA-SafeSec: Safety and security analysis for cyber-physical systems	Friedberg et al. (2017)	Power grid
Systems theoretic process analysis of information security: the case of aadhaar	Tarafdar and Bose (2019)	Digital Identity Program
Dual Governance: The intersection of centralized regulation and crowdsourced safety mechanisms for Generative AI	Ghosh and Lakshmi (2023)	Generative AI

Table 2.1.: Applications of STAMP and STPA in social and sociotechnical systems. The first 8 entries were retrieved from Patriarca et al. (2022).

- The safety control structure can reveal hazards in the interactions among hierarchical levels and/or actors, such as government, service providers, etc. that lead to harmful consequences to users. Additionally, responsibilities of safety measures can be assigned accordingly based on the safety control structure.

STPA to improve system safety often result in tangible and/or actionable system safety requirements. These deliverables can then be translated into safety designs and control actions.

However, although Leveson (2012) claims that STAMP and STPA is applicable to socio-technical systems, its application outside technical systems is very limited, which is shown in the Literature Review in Section 2.4. This research, hence, also evaluates the applicability of STAMP and STPA in socio-technical systems, particularly social media.

STAMP and STPA are used to answer RQ 3. The research first attempts to conceptualize the problem of unreflective habit formation on TikTok in STAMP terminology by phrasing the harms, losses, and hazards identified in RQ 2 in STAMP terminology. This problem conceptualization is explained in Section 7.1.4, both in text and conceptual model (Figure

7.3). Once the problem is conceptualized using this terminology, the research applies STPA method to analyze the problem and produces the following outputs:

- **System-level constraints**: directly inferred from the identified hazards.

- **Hierarchical control structure**: conceptualizes the roles of, and the feedback information received by, different actors in the regulating and operating TikTok in society.

- **Functional control diagram**: models the control actions of the system controllers, the feedback information sent to controlled processes and received by controllers, and the controlled process.

- **UCA**: highlights the causal factors of hazards in the system.

- **Controller constraints**: derived directly from the identified UCA and provide restrictions for controllers to prevent hazards.

- **Loss scenarios**: identify different situations where losses may occur.

3. Methodology

This chapter outlines the methodology used in this research, including data collection methods and the analytical approach taken to build the STAMP components.

Among the four philosophical worldviews outlined by Creswell (2008), this research aligns most closely with Pragmatism, which places understanding the research problem at the center of inquiry. Unlike Postpositivism, Constructivism, or the Participatory/Advocacy paradigm, Pragmatist researchers are not restricted to either quantitative or qualitative methods (Creswell 2008). Instead, they adopt methodological tools best suited to address the specific research problem. As Mahoney and Goertz (2006) explains, qualitative research seeks to understand the causes behind individual cases, while quantitative research aims to identify the average effect of causes across a broader population. Therefore, qualitative findings are typically more robust for understanding unique cases, whereas quantitative results provide stronger reliability for generalizing across larger samples.

The researcher chose a qualitative methodology with the goal of developing a deep understanding of individual user experiences. This choice reflects the belief that in-depth exploration of personal cases can uncover subtle but meaningful insights that may not be easily captured through a quantitative approach. However, this methodological decision comes with a limitation: the findings might not be generalized to a broader population beyond the studied individuals. Therefore, the credibility of the conclusions would benefit from future research using quantitative methods that arrive at similar results and can support broader generalization.

3.1. Desk Research

Desk research is used to principally answer RQ 1. First, to cover a broader concept of autonomy and avoid the shortcomings pointed out by Klenk and Hancock (2019), the focus of desk research is on the conceptualization of autonomy by looking into available scholarly literature about autonomy. Second, the focus is then shifted to the mechanisms employed by social media to enforce unreflective habit formation to answer RQ 1. Scholar literature explaining habit formations on social media is explored. The findings from the concept of autonomy and the unreflective habit formation mechanisms are synthesized to explain how unreflective habit formation is an expression of loss of user autonomy. This synthesis serves to justify the cause of the research, which is to address unreflective habit formation from the aspect of loss of user autonomy. In this way, it substantiates the discussions in RQ 2 and the STPA analysis in RQ 3.

Desk research supports answering RQ 2 in two main ways. Firstly, the synthesized answer of RQ 1 sets the focus of the interview questions to answer RQ 2a. Secondly, desk research is used to identify losses resulting from the identified unreflective habits found in RQ 2a. To

answer RQ 2b, the focus of the desk research is on the losses that social media use, through habits, can incur to users.

To answer RQ 3, desk research is used to map the control structure of the TikTok system at both the functional operation and the societal hierarchy levels. Ideally, interviews with experts - such as current or former TikTok employees, government staff - would provide the most valuable insights into the platform's internal operations and the government's management of the system. The interviews with TikTok employees could illuminate how features are conceived, prioritized, implemented, tested, and ultimately maintained or retired. Such detailed accounts would offer a comprehensive understanding of the decision-making processes that shape user experience. The interview with government staff would shed light on the current regulatory means, challenges and priorities of the Vietnam Government. However, due to time constraints, conducting these expert interviews was not feasible. As an alternative, the researcher employed desk research to collect the necessary data for mapping the mechanisms underlying TikTok's platform design and functionality, as well as the legislative governing of the system. The key areas of interest include:

- The different legislative documents the government use to manage social media.

- The individuals and entities at TikTok responsible for controlling and managing feature development.

This data was then used to build the functional control diagram and hierarchical control structure components of the STPA analysis. To construct the Functional Control Diagram, the researcher primarily consulted publicly available TikTok documents that describe its functionality and operations, such as job postings. Additionally, the researcher examined scholarly articles that attempt to analyze and critique TikTok's internal processes.

For the hierarchical control structure, the researcher expanded on the sources used for the control diagram by analyzing job listings from TikTok (TikTok 2025l), focusing on the TikTok T&S as this team is mainly responsible for ensuring user safety at TikTok (Keenan 2021). These listings were examined for information regarding job titles, team structures, and responsibilities. Additionally, the research reviewed official legislative documents, scholar articles and news about the government's measures to regulate social media in general.

In the context of desk research, confirmation bias - the tendency to focus on information that supports preexisting beliefs while disregarding contradictory evidence - can compromise the quality and objectivity of the data (Snyder 2019). To safeguard against such bias and maintain the integrity of the research, the researcher adhered to rigorous guidelines and applied clear validation criteria throughout the desk research process (Snyder 2019).

3.2. Interview

Interview is used to gather individual narratives to answer RQ 2a. Furthermore, it is used to build components of the STPA analsysis to answer RQ 3. This approach allows the researcher to follow diverse directions in the conversation related to the central research topic, unreflective habit formation. Although surveys can gather larger volumes of data, they constrain participants to a predetermined set of questions, limiting the depth of their responses. In contrast, interviews enable participants to articulate their motivations and experiences more fully, providing richer insights. Additionally, interviews offer the researcher the flexibility to

uncover unexpected themes that may not have emerged through literature review or desk research alone. For these reasons, interviews were chosen as the main data collection method to investigate the unreflective habits users develop while using TikTok.

Twelve Vietnamese female students from various colleges of Hue University participated in the study. The research applies simple random sampling. All participants were over the age of 18. Each interview lasted between 30 and 45 minutes. The interviews were semi-structured, allowing flexibility for the researcher to tailor follow-up questions based on participants' experiences. The researcher wanted to explore participants' rationality behind and emotions from their TikTok usage. Therefore, the questions followed this format: "When you see [something], what strikes your interest?", "What do you think about [something]?", "How does [something] make you feel?", "When you encounter [something], what do you do?", etc. These questions focused on various aspects of TikTok, such as usage context, content types, persona impressions and influence, feature usages, and the overlap between offline social circles and TikTok experience, etc.

Interviews, while valuable for exploring lived experiences, are susceptible to several forms of bias. Zahle (2024) lists out some common sources of bias as follows. One common source of bias arises when researchers inadvertently ask leading questions in an effort to elicit responses that align with their expectations or hypotheses (Zahle 2024). Emotional dynamics during the interview may further skew the interaction, as researchers' reactions - whether subtle or overt - can influence how participants respond (Zahle 2024). Additionally, a researcher's personal commitments or strong stances on an issue may cause them to misperceive or misrepresent participants' accounts, particularly when those accounts diverge sharply from the researcher's own views. This can result in a lack of descriptive adequacy in the data (Zahle 2024). Similarly, when researchers hold strong positive attitudes toward the research topic, they may unintentionally overlook expressions of dissatisfaction from participants, leading to a biased dataset (Zahle 2024).

Biases may also originate from the respondents themselves. One example is social desirability bias, where participants may withhold their true opinions and instead provide responses they believe are more socially acceptable or favorable to the interviewer (Choi and Pak 2004). Another example occurs when respondents attempt to infer the purpose of the research. In trying to be helpful or to please the researcher, they may tailor their answers to what they believe the researcher wants to hear, rather than providing genuine responses. This well-intentioned behavior can nonetheless distort the data and compromise the validity of the findings (Choi and Pak 2004).

3.3. Synthesis and Analysis

RQ 1 sets the focus for the interview questions used to answer RQ 2a. The unreflective habits identified in RQ 2a are then used to guide the desk research in identifying potential losses to answer RQ 2b.

The answer to RQ 3 comprises various components of the STPA analysis. (For detailed instructions of the analysis, please refer to the Handbook by Leveson and Thomas (2018)). The synthesis can be summarized as follows:

- The findings from RQ 2b lay the foundation for defining the purpose of the STPA analysis in terms of losses, hazards, and system constraints.

- Desk research is used to identify the control structure of the TikTok system.

- The mechanisms of habit formation by social media identified in RQ 1, the losses, hazards, and system constraints, and the control structure are used to identify UCA, controller constraints, and loss scenarios.

- Finally, the components of the STPA analysis are used to derive recommendations for relevant actors to mitigate the harms caused by unreflective habit formation.

4. Unreflective Habit Formation and the Loss of User Autonomy

4.1. General Accounts of Autonomy

Autonomy is generally understood as self-government or self-direction, that is, exercising one's own motives, reasons, or values (Stoljar 2013). The scholarly debates on the topic of autonomy center around the essential conditions to judge if an agent is autonomous (Stoljar 2013). In the *Stanford Encyclopedia of Philosophy*, agency either "denotes the performance of intentional actions" or "refers to the exercise of the capacity to perform intentional actions" (Schlosser 2019). The mainstream scholarly debates about agency center around the nature of intentionality of actions in agency. While these two concepts are closely related, and even used interchangeably in most cases, this research leans more on the concept of autonomy in exploring unreflective habit formation on social media. This is because intentionality is the result of decisions, which implies autonomy. In the case of unreflective habit formation, it is this intentionality that social media is tinkering with. The nature of unreflective habits is that they escape users' conscious awareness (Marin 2025). Social media are designed with "ludic loops" that reinforce compulsive engagement (Norlock 2021). Users are merely responding to designed rewards that social media provides to reinforce their behaviors (Esposito and Ferreira 2024). In this way, social media infringes on users' decision-making process of habit adoption. Therefore, the research explores unreflective habit formation on social media primarily under the lens of autonomy. Since autonomy is already a complex concept to unpack in itself, agency will not be dissected further than the definition by the *Stanford Encyclopedia of Philosophy* by Schlosser (2019).

Sahebi and Formosa (2022), in discussing the negative impacts of social media on autonomy, identified the three main types of autonomy in the context of social media: procedural, substantive and relational. Reviewing different theories, Stoljar (2013) summarizes that procedural autonomy is "achieved when the user undergoes, or has the capacity to undergo, an internal intellectual process of reflecting on her motivations, beliefs, and values, and then revising her preferences in the light of such reflection. This process is said to be 'content-neutral' because the outcomes of the process of critical reflection, whatever their content, will be autonomous." The procedural approach, therefore, evaluates the decision-making process to conclude if the user is autonomous. The nature of the decision itself, be it controversial or undisputed, harmful or beneficial, etc. is irrelevant in assessing the autonomy of a user.

Substantive theories define conditions for a user's choices to be considered autonomous. There are two subcategories of substantive autonomy, namely strong substantive and weak substantive (Stoljar 2013). Strong substantive theories correlate autonomy with a user's abiding by normative constraints (Stoljar 2013). Strong substantive theories, therefore, are seen as limiting autonomy because they only allow users a subset of choices that are morally

acceptable in the universal set of all possible choices (Sahebi and Formosa 2022). Furthermore, it is rather challenging to determine clear boundaries for subjective good against evil in our society. However, in cases of oppression, such as slavery, strong substantive theories advocate for social change, breaking the chains that tie people to their exploited positions. Therefore, strong substantive accounts of autonomy should not be neglected.

Weak substantive theories determine a user's autonomy by evaluating their positive attitudes toward themselves, typically self-respect, self-love, and self-esteem (Sahebi and Formosa 2022). As this approach is more lenient towards moral constraints and encourages positive self-attitudes, it is applicable in defining autonomy in the context of social media. So, this research will consider both sub-categories of substantive accounts of autonomy.

Relational theories acknowledge the fact that users are situated in densely connected social webs. Stoljar (2013) further divides this approach into causal and constitutive. The causal approach recognizes the exertion of external influences from both social relationships and socio-historical circumstances on a user's autonomy in two opposite directions: they either stimulate or stunt its development (Stoljar 2013). The constitutive approach, with a slightly diverging view, pinpoints the interpersonal or social conditions as the criteria for judging a user's autonomy (Stoljar 2013). This research will use the definition of causal relational autonomy in evaluating if a user's social relationships are constructive to or destructive of their autonomy.

4.2. User Autonomy in HCI

As a user interacts with TikTok, and more generally any social media platforms, they interact with what Gal (2018, p.8) coins as "Predicted Preferences Algorithms," where the algorithm "generates a simulation which attempts to mimic and predict consumer preferences" based on the available data that the algorithm has about both the consumer in question as well as the larger pool of consumers that use the same platform. Hence, the research will focus the concepts of autonomy in the context of HCI. The project will use "autonomy" and "user autonomy" interchangeably.

Conducting a systematic literature review in the field of HCI over 32 years (1990 to 2022), Bennett et al. (2023) identify several elements of user autonomy that scholars in the field of HCI focused on, some of which can be used to form the different dimensions of user autonomy that this project relies on.

The first element is self-causality and identity where user autonomy is described in terms of the users' causal involvement (in executing actions or making decisions) and/or in terms of users' self and values (Bennett et al. 2023). Self-causal is then further divided into 2 cases, namely executing (a user's direct engagement in tasks, processes, outcomes), and decision (a user's decision-making about personal choices, tasks, processes, and outcomes) (Bennett et al. 2023). The executing element aligns with the definition of agency by Schlosser (2019). The decision element loosely presupposes procedural account of autonomy, as it emphasizes the user's direct engagement in the process of reflective decision-making and executing. Using the procedural account of autonomy, the definition of agency by Schlosser (2019) and the definition by Bennett et al. (2023), autonomy in this element can be defined as:

- **Decision:** a user is autonomous if their decision is the result of a process of internal reflection

- **Execution:** a user is agentic if they can execute their intention

Regarding self-identity, there are two main cases, namely self-congruence (the degree to which outcomes align with a user's values and goals, without regard to their causal involvement) and self-construction (the impact on a user's identity, values, and goals) (Bennett et al. 2023). Self-identity, especially the case of self-congruence, discussions presuppose weak substantive account of autonomy, as it emphasizes the user's self-regard. Combining the weak substantive account of autonomy and the definition of self-congruence as defined by Bennett et al. (2023), autonomy in this element can be defined as:

- **Self-congruence:** a user is autonomous if their decisions and actions align with their values and goals, without regard to their decision-making process.

The second element is experience and materiality. Material autonomy refers to a user's enaction of their autonomy, whereas experiential autonomy denotes a user's subjective sense of agency when triggering an event (Bennett et al. 2023). This element embarks on a rather philosophical exploration, investigating the blurring boundary of (sub)consciousness. Despite posing intriguing intellectual challenges, this element is, unfortunately, out of the scope of this project.

The third element is independence and interdependence. Bennett et al. (2023) found that most HCI scholars equate independence, which is the user's ability to "act, think, and be free from the influence or control of others," with autonomy. On the other hand, interdependence takes into account the influence of social context in supporting or suppressing a user's autonomy (Bennett et al. 2023). This element is best understood using the lens of relational account of autonomy, especially the case of interdependence. Combining the relational account of autonomy and the definition of interdependence by Bennett et al. (2023), autonomy in this element can be defined as:

- **Interdependence:** whether a user's social relationships are constructive to or destructive of their autonomy.

Marin (2025) proposes the concept of Formative agency, which is built upon the concept of Situated agency. A situated account of agency recognizes that though a user is constrained by the environment they inhabit, they still retain a certain degree of agency in the way they respond to environmental stimuli (Marin 2025). Similarly, the concept of formative agency postulates that a user still retains their agency in how they transform themselves regardless of the desirability of the changing experiences (Marin 2025).

In analyzing the nature of the socio-technical environment (herein referred to as environment) created by social media platforms in fostering self-transformation for the formative user, Marin (2025) argues that there are three main ways an environment is hostile to the user's autonomy. Firstly, the environment makes it difficult for the user to choose among options by either opening up too few, too many or only necessary options. Secondly, the environment impedes long-term reflection on the user's values and goals or introduces hidden values. Thirdly, the environment locks in a user's identity by only showing them one facet of their self (Marin 2025). Marin (2025) argues that social media platforms prohibitively interfere with the formative agency of a user since social media are designed to limit the reflective capacities of the user, preventing them from making the conscious choice to adopt the habits or not before these habits have been ingrained into their own identities. Marin (2025) coins this process as unreflective habit formation on Social Media Platforms.

Under the strong account of autonomy, this hostile environment threatens user autonomy as it limits the perceived available options to them. The researcher refers to the idea of a hostile

environment introduced by Marin (2025), which is created by social media, to suggest that autonomy must be considered in the following dimension:

- **System hostility:** the way in which the socio-technical system impedes user autonomy

The term "system" is used instead of "environment" because the researcher intends for this dimension to refer specifically to the user-inspired designs and operations of social media systems. The word "environment" is less appropriate in this context, since in socio-technical systems it typically refers to the external context in which the system operates (Abbas and Michael 2023). Timms and Spurrett (2023) defines "hostile scaffolding" as the harmful implementation of technology that eases users' cognitive load at the cost of undermining their own interests, often to benefit other agents. Building on this definition, "System Hostility" refers to how user-inspired designs and operations of socio-technical systems undermine user autonomy and interests by influencing users' cognitive processes.

4.3. Loss of Autonomy and Social Media

Shelby et al. (2023) divide socio-technical harms at the micro-, meso-, and macro-levels by algorithmic systems into five different categories, among which is interpersonal harms. The harms falling under this category are those "intrapersonal elements through which people feel a diminished sense of self and agency" (Shelby et al. 2023, p.731).

Among the different harms belonging to this category, one is loss of agency or control, which occurs when "the use or abuse of algorithmic systems reduces autonomy" (Shelby et al. 2023, p.731). Shelby et al. (2023) argue that algorithmic systems threaten user autonomy by curating the information flow users receive using digital profiling. Sahebi and Formosa (2022) go to great lengths to explain the mechanisms with which algorithmic systems, or social media within the scope of this research, injure human autonomy. Firstly, preying upon a user's need for social contact and their fear of missing out, social media compel users to agree to give in to their usage data (Sahebi and Formosa 2022), which leads to digital profiling and, eventually, loss of agency and control. In other words, a user's interdependence has been utilized to prohibit their agency.

Secondly, capitalizing on users' attention, social media use personalized advertising to coax users to buy whatever advertisers want them to (Sahebi and Formosa 2022). In this case, even if users satisfy the executing and decision elements of autonomy, social media dupe them in confounding experiential and material autonomy, as Sahebi and Formosa (2022) argue that social media influence users beyond their conscious awareness. Thirdly, using design elements, social media can induce users to keep engaging with the platforms, i.e., control their behaviors (Sahebi and Formosa 2022).

Given the five dimensions of autonomy mentioned in Section 4.2, user autonomy can be examined by the following questions:

- **Decision:** To what extent do users go through critical internal reflection before making the decision when faced with the environmental stimuli from social media?

- **Execution:** To what extent do users intentionally execute an action when faced with the environmental stimuli from social media?

- **Self-congruence:** To what extent do the outcomes of users' actions and decisions align with their values and goals when faced with the environmental stimuli from social media?

- **Interdependence:** To what extent are users' social relationships constructive to or destructive of their autonomy when faced with the environmental stimuli from social media?

- **System hostility:** To what extent do social media systems impede user autonomy?

4.4. Unreflective Habit Formation on Social Media

Pinder et al. (2018), synthesizing various theories of behavioral models and HCI implementations related to human habit formation, defined habits as impulses to perform actions, acquired through repetition, and subconsciously triggered by specific contexts. Habitual behaviors can be carried out without requiring much conscious awareness (Pinder et al. 2018).

Based on Dual Process Theory - which accounts for the interaction between conscious and subconscious mental processes - there are two types of mental processes: Type 1 and Type 2 (Pinder et al. 2018). Type 1 processes are fast, non-conscious, cue-driven, and impulsive, whereas Type 2 processes are slow, conscious, goal-directed, and planned (Pinder et al. 2018). Using the Dual Process Theory, Pinder et al. (2018) explain the process of generating an action in response to contexts - illustrated in Figure 4.1, as follows:

1. Stable contexts - whether internal (bodily or mental states) or external (environmental stimuli) - provide cues, which are filtered by Type 1 or Type 2 perceptual processes.

2. These perceived cues trigger mental processes that generate potential response options. Type 1 processes produce impulsive responses from implicit memory, while Type 2 processes generate intentional responses from explicit memory. Regardless of type, the options are stored in the response stack.

3. Among the generated options, those that cross the "act threshold" will be enacted. In the case of conflicting or competing options, Type 2 processes are required for deliberate decision-making.

Pinder et al. (2018) argue that impulses are more likely to be enacted for two primary reasons. First, when decision-making between competing options is required, engaging Type 2 processes increases cognitive load. If this cognitive load is impaired - due to ego depletion, time pressure, limited working memory, or low self-control - individuals are more likely to default to impulsive responses. Second, impulses are generated more quickly than intentions, making them more likely to be enacted even before intentions are fully formed. As these impulses are repeatedly enacted in response to contextual cues, they become increasingly likely to surpass the act threshold in future scenarios.

As the definition suggests, there are three key components of habits: stable context, repetition, and impulse. Bridging Dual Process Theory, Modern Habit Theory, and Goal Setting Theory, Pinder et al. (2018) define the habit formation process as follows. Stable contexts provide cues which, if successfully processed through perception, trigger corresponding mental processes. These mental processes then generate potential response options. When
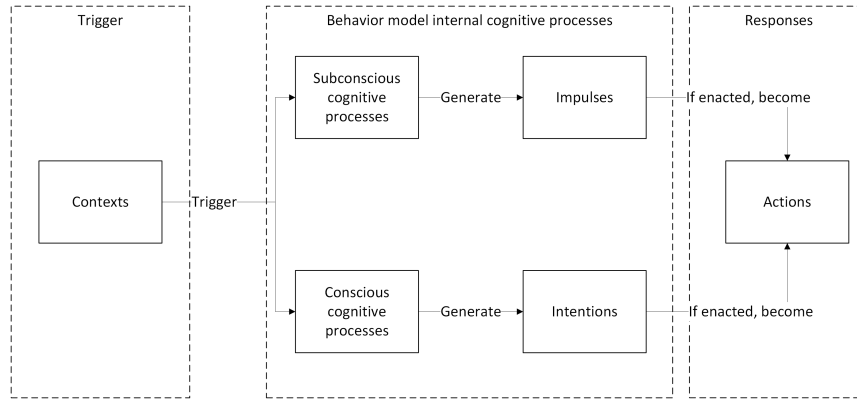
Figure 4.1.: The Action Generation Model adapted from Pinder et al. (2018).

an action is repeatedly enacted within a given context, it begins to require less cognitive effort to generate, increasing the likelihood that it will be selected as the default response in that context in the future.

Rewards, whether explicit or implicit, can accelerate habit formation (Pinder et al. 2018). Rewards play a crucial role in motivating non-frequent social media users to repeat certain actions on social media, thereby reinforcing those behaviors into habits (Anderson and Wood 2023). Once ingrained, these rewarded behaviors are carried out automatically (Anderson and Wood 2023).

While habits can result from conscious as well as subconscious mental processes, the danger of unreflective habit formation lies in its unconscious nature, which means that users are unaware of the process and therefore cannot critically reflect on or intentionally direct it (Marin 2025).

The design philosophy of social media is heavily inspired by the book "Hooked: How to Build Habit-Forming Products", which outlines how to build engagement by tapping into users' neurological reward cycles (Esposito and Ferreira 2024). According to the Hook Model, users experience internal or external triggers that prompt them to act. These actions are rewarded with transient satisfaction. To deepen the engagement, the system introduces variability - unpredictable outcomes that intensify desire, impair rationality, and reinforce behavior through variable rewards (Esposito and Ferreira 2024). Figure 4.2 illustrates the alignment between this cycle and the brain's neurological reward mechanisms as illustrated by Esposito and Ferreira (2024).

Examples of features that implement this model include infinite scroll, autoplay, push notifications, pop-ups, likes, emoji reactions, and view counts. These interface elements exploit cognitive vulnerabilities, impairing users' self-control and promoting reflexive, irrational behaviors that maximize engagement (Esposito and Ferreira 2024). More critically, because these features are designed to function organically, users' compulsive behaviors appear voluntary to them and, thus, they may not recognize these behaviors as externally induced (Esposito and Ferreira 2024).

Using the habit formation principles summarized by Pinder et al. (2018) and the mechanisms of habit formation on social media identified by Esposito and Ferreira (2024), the process of unreflective habit formation on social media can be explained as follows. The internal and
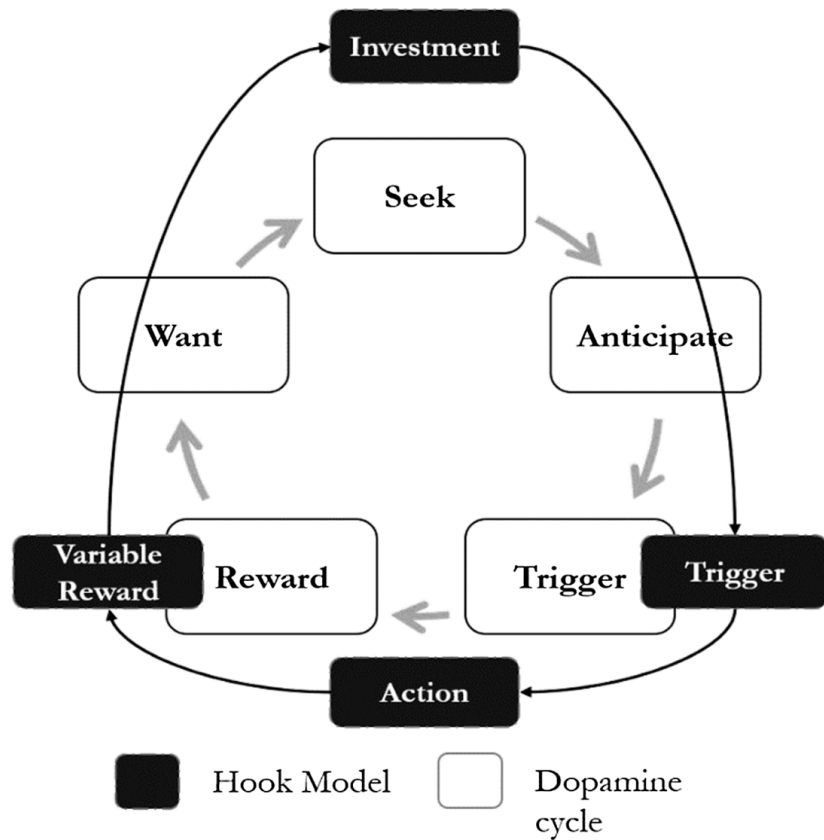
Figure 4.2.: The Hook Model as illustrated by Esposito and Ferreira (2024).

external stimuli that users experience constitute a context that can serve to trigger certain actions. To encourage users to carry out these actions, social media platforms are designed with features that offer rewards—whose unpredictability intensifies desire, impairs rationality, and reinforces behavior subconsciously. As users repeatedly perform these actions in response to the same contextual cues, their behavior becomes automated, requiring minimal cognitive effort. In this way, habits are formed. Figure 4.3 demonstrates this process.

In summary, social media features are built based on the functionality of the neurological reward cycle. The design of these features reduces users' cognitive ability and make them spontaneously seek for the rewards that social media provide. Users, however, are under the impression that these are their intentional actions. With time, these actions consolidate into habits. This is how users pick up unreflective habits of using social media.

Building on Esposito and Ferreira (2024), and referring back to the five dimensions of autonomy discussed in Section 4.3, users experience a loss of autonomy during unreflective habit formation in the following ways:

- **Decision**: Actions are performed automatically or with the illusion of agency. Without critical reflection, it is difficult to regard the user's intention of using social media as autonomous.
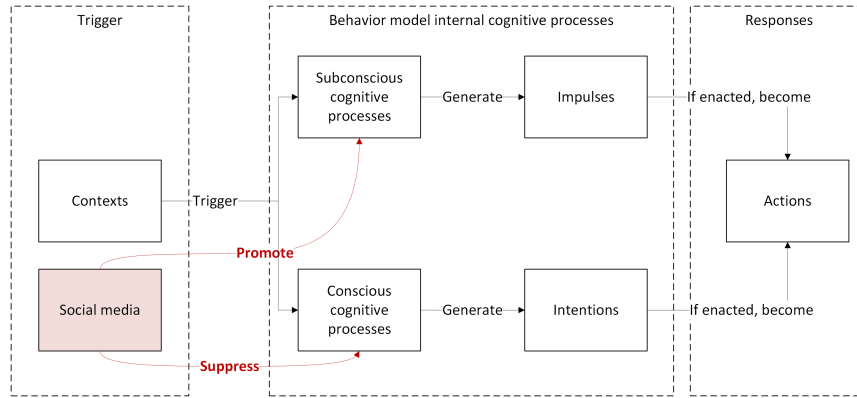
Figure 4.3.: The role of social media in the the Action Generation Model. The Action Generation Model is adapted from Pinder et al. (2018).

- **Execution**: Given the compromised decision-making process, the execution of those actions is also undermined.

- **System hostility**: The platform is deliberately designed to suppress rational evaluation and enhance impulsive behavior, thereby structurally limiting autonomy.

- **Self-congruence:** As habits gradually shape the self (Marin 2025), self-congruence is relevant in unreflective habit formation. However, whether unreflective habits are congruent with the self or not is a case-by-case and not a universal answer.

- **Interdependence:** While the aforementioned mechanism provides details on how the User Interface (UI) features shape habits, it doesn't explain the influence of the user's social circle on this habit formation. Therefore, the answer for this aspect is also inconclusive.

5. TikTok and Autonomy Infringements – Patterns from Interviews

Twelve Vietnamese female students from various colleges within Hue University participated in the study. The research applies simple random sampling. All participants were above the age of 18. Each interview lasted between 30 and 45 minutes and was conducted in April 2025.

Thematic coding was employed to identify emerging patterns within the semi-structured interviews. Since the interview format allowed for flexibility in how questions were posed and how participants responded, there was a variability in data. Even when participants conveyed similar messages, they did so using different language and framing. To uncover the common patterns among this variability, a coding scheme was developed. The conditions to satisfy the retainment of autonomy are debated among philosophy scholars in the field, making it a rather complex concept to investigate, as shown in Section 4.2. Furthermore, as mentioned in Section 3.2, the interview questions do not target users' perception of autonomy. Rather, they center around participants' observations, perceptions, thought processes, and emotions. Therefore, the researcher did not set the aim to find specific expressions of loss of autonomy in the encoding of the interviews. Instead, the researcher looked for and encoded the commonalities that emerged from different interview narratives. The researcher then compared the emerging patterns to the different dimensions of autonomy identified in Section 4.3 (decision, execution, self-congruence, interdependence, system hostility) and the mechanism of unreflective habit formation on social media in Section 4.4 (Hook Model) to identify unreflective habits to answer RQ 2a. The researcher encoded the process in this way to avoid cherry-picking evidence supporting the traits of loss of user autonomy and disregarding evidence that proves otherwise.

The researcher first identified common themes emerging from all interviews. These themes were then encoded as code groups. For each code group, the researcher identified more specific details, each of which was then encoded as a code. There are 92 such codes in total. The codes are shown in Table A.1 in Appendix A, which lists the full set of code names, descriptions, and the code group.

Not all themes and codes were equally relevant to the research objectives. After a comprehensive review of the data and reflection on the goals of the study, the researcher summarize the following key findings.

5.1. Using TikTok is an Automated Response to Stimuli

Participants indicated that they use TikTok primarily during free time - for example, between classes or during study breaks. Many mentioned using it to unwind, especially when bored or stressed or exhausted. To most participants, they picked up their phone and opened

TikTok whenever they felt bored. Most usage occurred at home after school, dinner, or chores. A significant number of participants used TikTok before they fell asleep. This is often described as a nighttime habit. Coffee shops emerged as common venues for filming TikTok videos with friends.

Q: *When do you normally use TikTok?*

A: *I think I use TikTok everywhere, as long as there is wifi. For instance, at school, I use TikTok during break. When I am too stressed with homework, I use TikTok. When I lie in my bed before I fall asleep, I use TikTok to see if there is anything new going on in there.*
— Interview 4

A: *"I normally use TikTok whenever I use my phone. When I use my phone, I mostly use TikTok. I don't message my friends or use Facebook that often. So, yeah, I mostly use TikTok whenever I use my phone. So, I often use TikTok around noon or at night, between 10 to 11pm. I also use TikTok whenever I have free time. I use TikTok pretty much everywhere, such as my workplace around noon and during my free time, but mostly at home at night."*
— Interview 6

	1 int_1.docx	2 int_2.docx	3 int_3.docx	4 int_4.docx	5 int_5.docx	6 int_6.docx	7 int_7.docx	8 int_8.docx	9 int_9.docx	10 int_10.docx	11 int_11.docx	12 int_12.docx	Totals
	26	45	26	23	25	32	31	41	17	26	15	35	
At Coffee Shops (8)		1	1	1			1	1			1		8
At Home (13)	1	1	1		1		2	1		1	1	1	13
Before Bed (10)	1	1	1			1		2		1		1	10
Between Classes/Study... (9)	1	1		1		1	3		1		1		9
During Free Time (15)	2	1		1	2	1	2	2	2	1		1	16
Totals	5	5	3	5	3	5	8	6	5	3	4	3	55
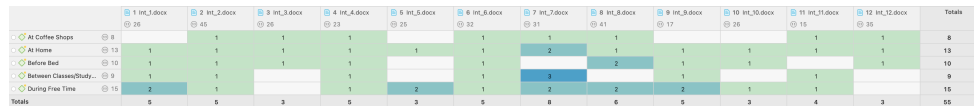
Figure 5.1.: Usage Context

In addition to Entertaining and Relaxation, participants cited several motivations for using TikTok, namely:

- Knowledge and Skills: TikTok is also used to learn new skills they can't gain from school, gain knowledge related to their studies or career orientations. For instance, one participant finds TikTok useful as she can seek for insightful knowledge aligning with her studies and interests in marketing. Another says that she learns skills related to practical language in tourism, which she is not taught at school. Participants find contents about knowledge and skills insightful, in-depth and practically useful.

- Information: Most participants get news update through TikTok. They say that TikTok is the fastest and most comprehensive news channel they can find. Most participants are interested in most recent and talked-about scandals, which they refer to as "dramas" out of personal curiosity and for in-person gossiping with friends. They don't seem to see any value in catching up with "dramas". They both are incidentally exposed to this type of news and actively seeking for it when they use TikTok. Additionally, they also see local news posts about crimes, traffic, weather phenomena, etc.

- Social Connection: Most participants use it to maintain their social connections, especially with friends. Common online activities include sharing entertaining or news videos, interacting with one another's posts to make the post owners happy, tagging one another in comments. Keeping streaks is seen by many as a way to strengthen their friendship. Common offline activities include gossiping about latest scandals or creating TikTok videos together.

- Online Shopping: Most participants find appearance and styling inspirations from Tik-Tok. Participants admire creators' beauty and style and seek to learn how to improve their own appearance through fashion and beauty routine tips. They might want to buy products or copy dressing styles to look beautiful like the creators. After finding suitable products, they make online purchases on TikTok Shop.

	1 Int_1.docx	2 Int_2.docx	3 Int_3.docx	4 Int_4.docx	5 Int_5.docx	6 Int_6.docx	7 Int_7.docx	8 Int_8.docx	9 Int_9.docx	10 Int_10.docx	11 Int_11.docx	12 Int_12.docx	Totals
	26	45	26	23	25	32	31	41	17	26	15	35	
Boredom Relief 6				2				1	1		2		6
Challenges/Tr... 19	1	5	6		1	3	2	1					19
Creating Vide... 23		4	6	1		5	1	2		2	1	1	23
Drama/Viral E... 24		2	6	4	1		1	6		3	1		24
Educational... 14			2	1			5	5		1			14
Entertainmen... 14	1	1			1			2	4	1	3	1	14
Enthusiastic... 7				2	1		2	2					7
Feeling conn... 12			2	2	3	1			2			2	12
Following Tre... 17		5	5	2		1			4				17
Funny/Entert... 20	2	3	2		2		1	2	2	2	1	3	20
Learning and... 10			2	1		1	1	4		1			10
Relaxation an... 4			1					1	1		1		4
Social Conne... 4			1	1		1			1				4
Shoppin... 13 48	2	7	2	2	1	3	8	5	5	4		9	48
Totals	6	27	35	16	11	16	22	35	15	16	7	16	222

Figure 5.2.: Motivations for Using TikTok

The usage patterns identified from the interviews suggest that using TikTok has become an automated, habitual behavior for many participants. This behavior is often triggered by various stimuli, including boredom, mental fatigue, stress, or periods of unstructured free time, particularly in relaxing settings at home. In response to these triggers, participants turn to TikTok in search of entertainment, which the platform readily provides through an abundance of content and opportunities for virtual social interaction. As this pattern of use is repeated, the act of opening TikTok in response to such stimuli becomes habitual.

It is important to highlight that mental and emotional exhaustion is a critical condition that facilitates habit formation. Under such conditions, users may experience cognitive impairments (Jonsdottir et al. 2013), making them more likely to default to habitual behaviors rather than deliberate, reflective actions (Porcelli and Delgado 2017). The nature of these triggering stimuli suggests that little, if any, cognitive effort is invested in the decision-making process prior to initiating TikTok use. As a result, the decision dimension of user autonomy is not fulfilled.

One might argue that among the motivations cited for using TikTok, acquiring knowledge and skills entails both serious cognitive effort and intentionality - thus fulfilling the dimensions of decision and execution. From this perspective, it would be incorrect to generalize TikTok use as purely compulsive. However, knowledge and skill acquisition inherently demands deliberate effort and mental energy, which users often lack when they open the app in search of relaxation or entertainment since they are stressed or tired. The argument for unreflective habit formation specifically targets the types of stimuli - such as mental fatigue or boredom that suppress rationality and promote habitual engagement. While the researcher encourages users to utilize TikTok for productive learning, they raise concerns that the platform capitalizes on moments of cognitive vulnerability to foster unreflective habits.

5.2. Underutilized Safety Settings Amid Popularity of Data Collection Features

Many participants were unaware of built-in TikTok features like keyword filters, sleep reminders, and screen time limits. One participant had used keyword filtering to avoid certain

content, while another had set a screen time limit but routinely bypassed it.

Participants took various measures to manage their privacy, such as making their accounts private, controlling visibility settings, and disabling comments. One participant configured their settings to conceal profile visit activity, and others restricted who could engage with their posts.

The "save for later" feature was frequently used to bookmark videos related to beauty products, fashion items, or dance trends. This allows them to retrieve videos later, especially those containing clothing or beauty product information, dancing trends they want to learn, or videos they find useful. Some save videos to retrieve product information for future purchase or to compare products.

Participants also used the "heart" feature to show appreciation and support creators. Participants heart videos they really like and find meaningful, or those featuring products they are interested in. They also heart videos to potentially retrieve them later. Hearting videos is an expected practice among friends to increase the interaction rate of each other's videos.

The use of visual filters was widespread. Participants used them to enhance their appearance, lighten makeup, or create funny effects in creating TikTok videos with friends. They commonly used the search bar to look up music, news, study resources, fashion items, or specific trends they want to make videos. Skipping videos - especially repetitive ones or those showing revealing outfits - was a common tactic to manage their feed.

> **Q:** *What other TikTok settings do you use, such as screen time limit or sleep reminder?*
>
> **A:** *I am aware of these settings. I use the screen time limit to set a limit of one hour. So, when I use TikTok past that limit, TikTok will notify me, requesting me to fill in my password. I simply fill in my password and continue using TikTok - especially when I am too absorbed in using TikTok, or when I don't have anything else to do, I just can't stop using it. I am not aware of the sleep reminder.*
> — Interview 3

	1 Int_1.docx	2 Int_2.docx	3 Int_3.docx	4 Int_4.docx	5 Int_5.docx	6 Int_6.docx	7 Int_7.docx	8 Int_8.docx	9 Int_9.docx	10 Int_10.docx	11 Int_11.docx	12 Int_12.docx	Totals
	26	45	26	23	25	32	31	41	17	26	15	35	
Filters 11	1		1	2						1	1		11
Liking (Hearting) Videos 11	1	1		1	1	1	1	2		2	1		11
Limited Awareness of... 19	3	3	1		3	1	2	1		2		1	19
Privacy Settings 8	1	1	1			1		1			1	1	8
Saving Videos 22	1	4	3	1		4	2	2	2			3	22
Totals	7	9	6	7	3	9	6	7	2	5	4	6	71

Figure 5.3.: Awareness and Usage of Common and Safety Features

Safety features on TikTok, particularly screen time management tools, appear to function more as symbolic gestures than as effective safeguards. Although TikTok is aware of the limited effectiveness of these features, it continues to promote them as evidence of its commitment to user safety, primarily to appease policymakers (Haidt 2023). The design of these tools - hidden within layers of settings and disabled by default - further limits their usefulness. As users tend not to explore or adjust system settings in detail (Mik 2016), these features remain largely ineffective.

In contrast, features that facilitate user profiling and content personalization, such as the "like" and "save for later" buttons (TikTok 2025i), are well received by users. Participants appreciate these features for helping them access preferred content more efficiently. However, these mechanisms raise concerns regarding user autonomy, particularly in the decision and execution dimensions. TikTok does not allow users to configure what data is collected or

how it is used in content recommendations. While users actively "like" content, they do not explicitly consent to have those actions serve as input for TikTok's recommender algorithms. Therefore, the execution dimension is compromised, as the data collection and usage for content personalization is not intentionally authorized. Moreover, even if users are aware of the data implications of "liking" a post, their need for social connection often compels them to engage, effectively trading autonomy for social connection (Sahebi and Formosa 2022). In this sense, the decision dimension is also undermined, as choices are made under social pressure rather than genuine self-governance.

5.3. Preference for Self-Identified Personas

Most participants identified with the image of a simple and elegant woman, which they perceived as socially appropriate. They frequently encountered posts featuring TikTokers whose personas aligned with their own - often described as "cute." These TikTokers served as sources of inspiration, motivating participants to engage in online purchases (see Section 5.4) and to create videos collaboratively with friends (see Section 5.5). When encountering content featuring sexually suggestive appearances, participants commonly reported feelings of repulsion and tended to skip such videos. Consequently, the content they consumed was predominantly aligned with their own self-image, rarely featuring TikTokers with more overtly sexualized presentations.

> **Q:** *Why do you skip these videos where women dress too revealingly?*
>
> **A:** *"I think these creators must post these videos purposefully. Personally, I think these videos look repulsive, and it depreciates their own values. I think I am quite conservative. I don't like to show too much of my bare skin, especially on social media because there are so many people viewing these contents. Anyway, since everyone is free to upload whatever they want, everyone is also free to judge."*
> — Interview 8

Participants' responses indicate that they actively assert their autonomy across the dimensions of decision, execution, and self-congruence. For instance, upon recognizing their aversion to the sexually suggestive female image, they consciously choose to reject such content (decision) by skipping these videos when they appear on their feed (execution). As a result, their TikTok experience is curated around creators whose personas align with their self-identified image (self-congruence). However, as the algorithm adapts to this preference and continually surfaces content featuring only "cute-looking" women, it reinforces a singular aspect of their identity. This algorithmic curation restricts the diversity of persona expressions available for exploration and development, thereby creating a system that is potentially hostile to self-transformation (system hostility).

5.4. Rationality Over Compulsiveness in Shopping Behaviors Conceals Unchecked Desire Induction

> **Q:** *Can you describe the entire process you go through from the time you see a product until the time you buy it?*

> **A:** *After viewing the entire video, I will search for more information on the product. Take the case of eye coloring powder for instance. In the video, the creator will introduce the product and apply the product directly on their eyes. Then I will read the comments of the other users. They will request for comparison between this featured product and other products that they like and are using. Other users also ask the compatibility of this product on different skin types.*
> — Interview 7

Nearly all participants reported making purchases via TikTok. Several recurring themes were observed:

- Beauty Aspirations: Participants were drawn to aesthetically pleasing creators and aimed to replicate their appearance - but only if the style aligned with their own.

- Product Fit: Items needed to suit their body or skin type, or they would not proceed.

- Community Feedback: Comments from others helped participants evaluate product suitability.

- Search Functionality: Search tools, including image-based search, helped them locate products.

- Content Repetition: They were aware that TikTok nudges them to buy products by repeatedly showing them these products. However, the repetition was irritating to them because they no longer want to see contents of similar products. Besides, whenever they decide to buy a product, either immediately or later, they save the videos or add the product to their basket.

Participants did not make impulsive purchases. Their decision-making process typically included:

- Engage with the post: Viewing, reacting, saving, and reading comments to assess compatibility with their needs before adding the item to their basket. Participants shared videos with friends either upon request or to coordinate matching outfits.

- Access product links: Using links in promotional posts, which led to TikTok Shop or external e-commerce platforms.

- Gather information: Reviewing comments both on the video and the product page to assess product fit.

- Product comparison: Comparing different sellers or platforms to find the best deal.

- Reflection: Taking time to consider whether to proceed with the purchase.

- Financial assessment: Reviewing monthly finances to decide whether a purchase is feasible.

- Savings: Using money saved from part-time jobs to fund desired purchases.

At first glance, it may seem paranoia to argue that user autonomy is compromised when users make purchases on TikTok, as multiple dimensions of autonomy is substantially exercised. Users often engage in deliberate evaluation - reading product reviews to assess suitability, fit, and quality - demonstrating autonomous decision-making. The act of purchasing is a direct result of this cognitive deliberation, fulfilling the execution dimension. Furthermore, the products they choose often align with their personal interests and aesthetic preferences, satisfying the criterion of self-congruence.

	Admiration fo... 21	Comparing pr... 13	Considering p... 13	Desire 19	Discovering P... 18	Intrigue 8	Reading Revie... 7
Admiration for appeara... 21				12 (0.43)	7 (0.22)	3 (0.12)	
Comparing products an... 13			5 (0.24)	2 (0.07)	2 (0.07)	1 (0.05)	3 (0.18)
Considering product fit 13		5 (0.24)		2 (0.07)	3 (0.11)	2 (0.11)	7 (0.54)
Desire 19	12 (0.43)	2 (0.07)	2 (0.07)		14 (0.61)	7 (0.35)	2 (0.08)
Discovering Products T... 18	7 (0.22)	2 (0.07)	3 (0.11)	14 (0.61)		9 (0.53)	4 (0.19)
Intrigue 8	3 (0.12)	1 (0.05)	2 (0.11)	7 (0.35)	9 (0.53)		3 (0.25)
Reading Reviews and F... 7		3 (0.18)	7 (0.54)	2 (0.08)	4 (0.19)	3 (0.25)	

Figure 5.4.: Shopping Behaviors

However, this surface-level analysis conceal a more insidious dynamic. The manipulative mechanisms of social media frequently escapes users' conscious awareness, subtly directing them toward outcomes they perceive as self-determined (Malik et al. 2022; Esposito and Ferreira 2024). By repeatedly presenting images of women that mirror users' idealized personas, TikTok instills in users a desire to emulate those figures. This desire is rendered instantly actionable through seamless advertisements for products that promise to fulfill it. While users may recognize the commercial intent behind these tactics, they often remain unaware of the deeper manipulation, which is that TikTok has strategically cultivated their desire to appear attractive. Interview narratives suggest that while participants exhibit skepticism toward overt commercial messages, they are nonetheless eager and enthusiastic in their admiration for and aspiration to resemble the idealized women featured on the platform. User autonomy, hence, is infringed to a large extent in the case of online shopping on TikTok.

5.5. Creating Videos with Friends: Everyone Can Look Cute on TikTok

Participants enjoyed creating videos with friends to post on TikTok. They often mimic two main types of content on TikTok. One popular genre involves trendy, cute hand choreography set to popular songs as soundtracks. These videos are often playful, visually engaging, and easy to replicate, making them ideal for mimicking. Another common content type includes humorous interactions using face-distorting filters, which participants find entertaining both to create and to watch. These funny, exaggerated expressions is often seen as a source of lighthearted amusement among friends. Video creation often began with trend discovery. Participants saved, shared, and rehearsed content with friends, then filmed - often at coffee shops.

> **Q:** *What do you see when you look at trending dancing TikTok videos?*
>
> **A:** *Whenever I see a dancing trend on TikTok, it seems to have reached a large audience. So, if I look at 10 videos, I will see around 8 videos of people doing that trend. Then, I start to feel FOMO, and I really want to cover or copy the trend to make TikTok videos. Most of the time, the trends that I follow are videos where people make dancing movements with their hands, not their entire body.*
> — Interview 3

Participants were motivated to make TikTok videos for a variety of reasons. Some want to engage in something trendy and enjoy the process with friends, treating it as a fun, social bonding activity. Others view content creation as an opportunity to practice marketing skills,

especially when promoting something meaningful to them, such as a business or event. In these cases, participants reported having more serious discussions to strategize and improve the chances of their video going viral. Social validation was a key motivator. There is an element of randomness regarding the virality of videos: one can never know for certain if their video will become viral. Disappointment with poorly performing videos sometimes led to deletion or improvement efforts.

	Challenges/Tr... 19	Creating Vide... 23	Discussing Ti... 9	Enjoying cute... 7	Saving Videos 22	Social Conne... 4
At Coffee Shops 8	0 (0.00)	3 (0.11)	1 (0.06)		0 (0.00)	1 (0.09)
Challenges/Trends 19		9 (0.27)		3 (0.13)	3 (0.08)	
Creating Videos Togeth... 23	9 (0.27)		2 (0.07)	5 (0.20)	5 (0.12)	2 (0.08)
Discussing TikTok Cont... 9		2 (0.07)		1 (0.07)	0 (0.00)	1 (0.08)
Enjoying cuteness 7	3 (0.13)	5 (0.20)	1 (0.07)			1 (0.10)
Feeling connected with... 12	0 (0.00)	1 (0.03)		0 (0.00)	0 (0.00)	1 (0.07)
Feeling embarrassed a... 2	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Feeling FOMO 1	1 (0.05)	1 (0.04)		0 (0.00)	0 (0.00)	0 (0.00)
Filters 11		6 (0.21)				1 (0.07)
Following Trends and F... 17	6 (0.20)	12 (0.43)	2 (0.08)	5 (0.26)	4 (0.11)	2 (0.11)
Funny/Entertaining Vid... 20	1 (0.03)	5 (0.13)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Positive emotions with... 10	0 (0.00)	1 (0.03)	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.08)
Sharing Videos Directly 27	1 (0.02)	2 (0.04)	2 (0.06)	0 (0.00)	2 (0.04)	0 (0.00)
When Out with Friends 2	1 (0.05)	1 (0.04)	0 (0.00)	0 (0.00)	1 (0.04)	0 (0.00)

Figure 5.5.: Creating Videos

Similar to shopping behavior, the act of creating and posting TikTok videos with friends initially appears to be a fully autonomous activity. Users do not blindly replicate all the content they encounter; rather, they selectively create videos that align with their preferences (self-congruence) and are simple enough to emulate (decision). The production process itself - often involving rehearsals and coordination with peers - demonstrates deliberate planning and intentionality, fulfilling the execution dimension of autonomy.

Yet, as with shopping, the manipulative architecture of TikTok once again successfully escapes users' perception, crafting an illusion of user control. By consistently presenting videos featuring women who embody idealized personas, TikTok subtly fosters in users a desire to emulate those figures. This desire is made readily actionable through TikTok's built-in, highly effective editing tools. A few simple adjustments to lighting, color, or facial contours enable users to craft content that appears visually appealing with minimal effort. While users often express skepticism toward overt commercial advertisements, they readily embrace videos featuring conventionally cute and charming women. Their vigilance diminishes further in the face of TikTok's implicit promise: the possibility of virality. The prospect of achieving celebrity-like status at seemingly no cost is an enticing offer. Although not all videos achieve viral success, even modest levels of engagement - likes, comments, or views from friends - provide enough social gratification to encourage continued participation. In this way, TikTok entices users to engage ever more deeply, cultivating the perception that users are in control while subtly steering their behavior.

In this scenario, users' voluntary participation does not disprove the infringement on autonomy when evaluated through the lens of the system hostility dimension. The socio-technical system systematically restricts the range of meaningful choices available to users (Marin 2025). For instance, TikTok's recommender algorithms amplify popular content trends by inundating users' feeds with similar content. As a result, users face a false dichotomy: either join the trend to maintain relevance or opt out and risk social exclusion. In this constrained

system, users' autonomy is compromised - not because they are unaware of their actions, but because the system subtly coerces them into specific behavioral patterns under the guise of freedom.

5.6. Streak: When Lighthearted Virtual Social Connection Becomes a Duty

Nearly all participants reported maintaining streaks with friends. Many participants viewed streaks as a way to stay connected with their friends, especially when they cannot talk daily. Some even see it as strengthening their friendships. Additionally, participants view streaks as a "testament of our friendship" or even a romantic relationship, signifying the number of consecutive days of interaction. Videos shared for streaks typically featured funny or viral content.

> **Q:** *You mentioned fire streak. What is it?*
>
> **A:** *Every day, two people in the same individual chat room have to send some messages in the chat room. If only one person doesn't do that, we will lose the fire streak. If a fire streak is 100 days, it means that two people have been sending messages 100 days consecutively. If I were too busy that day to not log in to TikTok to send messages to my friends, we would lose that streak. That streak is like a testament of our friendship. If you are in a relationship, then that streak is a testament to our relationship. But recently, TikTok introduced a new feature that allows us to resume our streak. So, even if we don't send messages for one day, we can still resume the streak. So, I feel less pressured. Imagine you have a 1000-day streak with someone. But then you lose that streak, that person (either your friend or your romantic relationship partner) would grumble at me. So, I feel pressured to keep the streak. I normally share videos on TikTok simply just to maintain the streak. It's not about sharing videos to view anymore. Sometimes, I simply share some random videos just to maintain the streak. But if I come across some videos that I really like on TikTok, I will share it with my friends instantly.*
> — Interview 3

The perceived values of streaks drove some participants to maintain it with their friends. These participants enthusiastically thought that maintain streaks is very important. They even organized celebration or showing off their streaks milestones on social media. Those who did not share this sentiment did it to avoid social repercussion, i.e. avoiding complaints or grumbles from friends. One recalled incidents in the past when her friends called her to remind her to keep their streak. Regardless of motivation and commitment level, all participants find the ability to resume streaks a relief.

> **Q:** *What drives you to keep these streak? What if you don't keep the streak for 1 day?*
>
> **A:** *I share contents with my friends simply to respond to my friends' sharing. If I don't keep the streak even for one day, my friend will grumble at me when I meet them in class. Sometimes, I forget to log in to TikTok, so I don't send any message to my friends. Then, two days are deducted from our streak. When I meet them in class, they grumble at me for losing the streak. Sometimes, my friends even call me to remind me to keep the streak if they see that I am inactive or I don't log in to TikTok.*
> — Interview 5

		Maintaining "... 22
Boredom Relief	6	4 (0.17)
Creating Videos Togeth...	23	1 (0.02)
Discussing TikTok Cont...	9	2 (0.07)
Drama/Viral Events/News	22	3 (0.07)
During Free Time	15	1 (0.03)
Fashion and Clothing	11	1 (0.03)
Feeling connected with...	12	13 (0.62)
Feeling pressured abou...	5	5 (0.23)
Food–Related Content	11	2 (0.06)
Funny/Entertaining Vid...	20	4 (0.11)
Sharing Videos Directly	26	19 (0.66)

Figure 5.6.: Maintaining Streak

Through "streak," TikTok effectively transforms social pressure into a behavioral enforcer, strategically undermining user autonomy within the interdependence dimension. By capitalizing on users' intrinsic need for social connection, the platform induces users to ascribe significant symbolic value to streak, positioning them as milestones of relational closeness. Users express appreciation of the feature as a convenient reason (or excuse, rather) to exchange superficial messages with acquaintances. Beyond the fleeting excitement of perceived social bonding, the feature offers little substantive benefit other than a (self-)enforced duty to maintain it. Streak subtly impose an ongoing obligation: users feel compelled to uphold their streaks not only for themselves but also to avoid disappointing peers. In effect, users who harbor an attachment to streaks act as enforcers, applying pressure on their friends to maintain them. As the social circle is exploited to reinforce the behavior, users' autonomy is increasingly constrained. The interdependence dimension thus becomes limiting to user autonomy.

5.7. TikTok Can Read Minds

Participants have a strong perception that TikTok's algorithm personalizes their content feed based on their interactions and interests. Many believe that TikTok can "read their mind" and curate their feed based on the following interaction: 1) viewing the video long enough; 2) liking, saving or sharing a video; 3) searching for a particular concept, item, topic, trend, etc. Participants can vividly recall instances when this happened to them. Most participants notice that after interacting with promotional videos, they often see videos featuring similar products. They interpret this repetition as TikTok's attempts to nudge them to make a purchase. The duration of the repetition vary for different participants, ranging from one to several days.

Many participants appreciate that TikTok recommends contents matching with their interests. To one participant, it is the personalization that made them dread the prospect of their TikTok account being locked.

> **Q:** *What if your TikTok account is locked and you cannot retrieve it?*
>
> **A:** *I think I will be slightly sad. This is because as I have used it for quite some time, it knows my interests and preferences, so it can recommend to me the contents that I like to view. Now that I have to create a new account, TikTok will recommend everything to me, including the contents that I don't like. So, I find it quite annoying.*
> — Interview 9

On the other hand, the excessive repetition can make them feel fed up or bored. As an attempt to curb the contents on their feed, some participants actively skip videos they are not interested in or search for different topics. They all notice that these tactics work for them.

While participants recognize the personalization and its link to their activities on TikTok, they don't always understand the exact mechanisms. One participant assumes that Siri, in eavesdropping on her offline conversations, feed data about her interest to TikTok. Another believes that Shopee and TikTok are connected, saying that whenever she searches for something on Shopee, she instantly sees many videos featuring those products on TikTok.

Overall, participants have a clear sense that TikTok's algorithm actively curates their feed based on their interactions, leading to both appreciated personalization and sometimes frustrating repetition. They also recognize the algorithm's role in influencing their product discovery and potential purchasing decisions.

> **Q:** *How often do you see each of these content types?*
>
> **A:** *It depends. I think TikTok can read my mind. If I am interested in and view a food video, and then react heart to it, I see so many videos showing food later. If I see music videos, then more music videos will appear. If I simply skip some contents, then I will not see videos showing similar contents that often.*
>
> **Q:** *How do you know that TikTok can read your mind and know that it shows you contents similar to the videos that you are interested in?*
>
> **A:** *For instance, if I view videos about food, I see many videos about food later. If I view a mukbang video, the next video is about a person reviewing food. As long as it is about*

food, I will see it. And these videos are by different creators. For music, if there is a post about a video showing two pictures with a soundtrack, then the next will be about the singer, or the lyrics. The songs can be in both Vietnamese or foreign. Or, I will also see trendy songs. If I see videos where people are using CapCut, then I will see other videos where people use CapCut. All of these videos are by different creators. The more trendy a song is, the more videos using this song as the soundtrack will appear.

Q: *How does it make you feel?*

A: *I feel quite normal. This is because I am interested in such contents. So, it just appears based on my needs. These contents make me feel less stressed.*
— Interview 11

	Acceptance of Re... 11	Awareness of... 13	Being Influen... 7	Feeling fed-u... 14	Importance of... 4	Perception of... 40	Skipping Vide... 11
Acceptance of Repetition 11		1 (0.04)	2 (0.12)	5 (0.25)	2 (0.15)	8 (0.19)	2 (0.10)
Awareness of Promotio... 13	1 (0.04)		4 (0.25)	1 (0.04)		3 (0.06)	1 (0.04)
Being Influenced by Re... 7	2 (0.12)	4 (0.25)		1 (0.05)		6 (0.15)	
Feeling fed-up with rep... 14	5 (0.25)	1 (0.04)	1 (0.05)			11 (0.26)	5 (0.25)
Importance of Personal... 4	2 (0.15)					4 (0.10)	
Perception of Algorithm 40	8 (0.19)	3 (0.06)	6 (0.15)	11 (0.26)	4 (0.10)		6 (0.13)
Skipping Videos 11	2 (0.10)	1 (0.04)		5 (0.25)		6 (0.13)	

Figure 5.7.: Co-Occurrence Table of Perception of Algorithm

	1 Int_1.docx 27	2 Int_2.docx 47	3 Int_3.docx 27	4 Int_4.docx 24	5 Int_5.docx 25	6 Int_6.docx 32	7 Int_7.docx 31	8 Int_8.docx 41	9 Int_9.docx 17	10 Int_10.docx 26	11 Int_11.docx 15	12 Int_12.docx 35	Totals
Acceptance of R... 11						1	1	1	1	2	3	2	11
Awareness of Ad... 4						2		1			1		4
Being Influence... 7	1	2	2									2	7
Feeling fed-up... 14					2	2	2	4		3		1	14
Importance of P... 4								1	1	1	1		4
Perception of Al... 40	4		4	4	2	4	6	3	2	4	5	2	40
Skipping Videos 11						1	2	3	2	1	1	1	11
Totals	5	2	6	4	4	10	11	13	6	11	11	8	91

Figure 5.8.: Perception of Algorithms

The self-congruence dimension of autonomy appears to be maintained, as users are consistently presented with contents that align closely with their preferences. Despite the algorithm's influence on the curation of TikTok feeds, users actively engage in autonomy-preserving behaviors, such as skipping repetitive content or searching for specific videos, thereby exercising the decision and execution dimensions of autonomy. However, these coping mechanisms prove to be temporary and limited in efficacy. Once users exhibit new engagement patterns, the algorithm quickly recalibrates and redirects content exposure, creating a dynamic akin to a perpetual tug-of-war between user intent and algorithmic adaptation. Due to the algorithm's adaptability, users possess few actionable strategies to meaningfully influence their content feed. Thus, the execution dimension is constrained. More critically, TikTok's pervasive content curation, driven by opaque and forced data collection practices, compromises user intimacy by fostering a deceptive, altruistic human-machine relationship (Savolainen and Ruckenstein 2024). Consequently, the false intimate altruistic relationship makes it hard for users to critically assess TikTok's pervasive actions and, hence, infringes on the exercising of their autonomy (Savolainen and Ruckenstein 2024).

5.8. TikTok is Positively Received by Users

Many participants express generally neutral feelings towards TikTok. When rating their overall experience neutral, most of these participants express concerns about negative contents. Some of these participants enjoy using TikTok for entertaining, educational or informational purposes.

However, a significant number of participants also hold generally positive views. They see it as a source of entertainment and fun, a way to relax and destress, and a means of social connection and interaction with friends. Some find value in TikTok for learning and information, including news updates, product reviews, study tips, and gaining knowledge. They also appreciate TikTok for providing quick updates on news and trends.

Some participants also experience negative emotions when the app is inaccessible, feeling sad, lost, or unsure of what to do. When TikTok becomes inaccessible, some report feelings of sadness or even aimlessness, as if something essential is missing. They often resume usage at the earliest opportunity, either by viewing TikTok contents on their friends' device with them, or waiting for the app to function again. Some participants reported that when faced with such situation, they would use other social media, such as Facebook or Instagram, instead. Only one participant admitted to bordering on addiction - but even she rationalized it by saying others around her behaved the same way. No participant appeared to challenge their emotional attachment or reliance on the platform.

	Generally Neu... ⑨ 9	Generally Pos... ⑥ 6
Awareness of Addictive... ④ 4	1 (0.08)	0 (0.00)
Beauty and Makeup ⑮ 15	0 (0.00)	2 (0.11)
Concerns about Negati... ⑤ 5	8 (1.33) ⊙	2 (0.22)
Educational/Informatio... ⑭ 14	2 (0.10)	4 (0.25)
Enthusiastic with absor... ⑦ 7	1 (0.07)	8 (1.60) ⊙
Fashion and Clothing ⑪ 11	0 (0.00)	2 (0.13)
Funny/Entertaining Vid... ⑳ 20	4 (0.16)	3 (0.13)
Inspiration and Aspirati... ⑮ 15	2 (0.09)	0 (0.00)
Learning and Information ⑩ 10	4 (0.27)	6 (0.60)
Value for Information a... ④ 4	1 (0.08)	4 (0.67)

Figure 5.9.: Opinions of TikTok

Q: *What if your phone is broken so that you cannot surf TikTok at the times when you normally do, such as at night?*

A: *"I will definitely feel something is missing. I will try to figure out other ways to surf TikTok, such as using my laptop. I think it is like an addiction. I can use the laptop for*

a while to surf TikTok, but eventually I want to get my phone back to use TikTok on my phone. I can't bear it if I cannot surf TikTok. It's like an addiction. I notice that many people share this sentiment. It's like everyone must use TikTok. They are not accustomed to not surfing TikTok. All the people around me surf TikTok."
— Interview 6

Q: *What if your phone is not broken, but TikTok cannot be opened on your phone?*

A: *I will simply surf another social media. Many times in the past, my TikTok app was broken so I couldn't use it. I resorted to surfing Facebook. I personally felt disappointed and let down by the app. I cursed the app for not functioning when I needed to use it during my short study breaks. Then, I used other social media platforms. I could use it later when I went home.*
— Interview 7

TikTok has become an integral part of users' daily routines, so much so that it is often perceived as indispensable. This finding illustrates the platform's deep entrenchment in users' lives. It is difficult to hold users at fault for their ready acceptance of what may be seen as a manipulative force cloaked in benevolence. TikTok provides them with comfort during moments of boredom, mental fatigue, and stress, curating content that aligns with their emotional state. It keeps them informed about current events and social trends, and introduces them to aspirational figures whose appearances and lifestyles they admire. Moreover, not just stop at showing, TikTok supports their aspiration by promoting products that promise to help users emulate the personas they aspire to. Through accessible visual editing tools, users can craft aesthetically pleasing content that receive social validation in the form of views, likes, and shares. TikTok also fuels their hope of going viral from their posts.

Although users may be aware of certain flaws within the platform, these are often overshadowed by the perceived benefits TikTok offers - convenient entertainment, self-expression, and social connection, all at no financial cost. As a result, users often fail to critically evaluate their engagement with the platform. In this way, TikTok hinders user autonomy by ingraining habits while limiting user critical reflective capacity.

6. Discussion

6.1. Unreflective Habit Formation on Social Media and Losses to Users

The key characteristics of unreflective habit formation are: (1) repeated actions driven by rewards, and (2) a lack of critical reflection or intentional control over the habit's development (Marin 2025). Based on these criteria, the behaviors exhibited by participants in this study suggest that their use of TikTok constitutes an unreflective habit.

As discussed in Section 5.1, participants habitually use TikTok when they are bored, mentally fatigued, or have spare time and want to seek entertainment. Based on the habit formation model by Pinder et al. (2018), these conditions form the context that triggers the action-generation process. Stress has been shown to impair cognitive functioning (Jonsdottir et al. 2013), reduce goal-directed behavior, and drive individuals toward automatic, habitual actions (Porcelli and Delgado 2017). In such contexts, where users have a preference for low rather than intensive cognitive loads, impulses are more likely to dominate over intentions, as they place less strain on cognitive resources (Pinder et al. 2018). Therefore, using TikTok in these situations is more likely to be a heuristic impulse than a rational intention.

As participants use TikTok, they are rewarded with entertaining or relaxing content. The platform is intentionally designed with features that reward users for their time and attention (Esposito and Ferreira 2024). These rewards make the act of using TikTok more compelling - and thus more likely to be repeated - when users encounter similar contexts in the future. Over time, as this behavior becomes the default response, it solidifies into an impulsive habit. Because the processes of action generation and enactment are driven by impulse, this constitutes a case of unreflective habit formation.

Three major losses arise from unreflective habit formation. First, users experience a loss of time. While one could argue that using TikTok in one's free time is a legitimate form of relaxation, this assumes users can control their engagement, which is not the case in reality. This is because social media user interface lacks effective stopping cues (Lewis 2017; Esposito and Ferreira 2024). When participants use TikTok during the 5-minute break between two classes at school, there is a stopping cue for them: they must stop using to resume the next class. However, most participants report often use TikTok at home - an environment with no external constraints. TikTok offers built-in tools to support time management, but most participants are unaware of them or ignore them. This finding is consistent with Mik (2016), who argues that such features are often deactivated by default and underutilized by users with low digital literacy. One participant reported using screen time limits, but admitted to bypassing them when they were heavily immersed in surfing TikTok. Consequently, TikTok usage often consumes users' time in ways that are beyond their control, representing a costly loss of an irreplaceable resource.

Second, users risk losing their ability to concentrate for extended periods. There are two main causes for this. First is continuous partial attention, which refers to the state of constantly scanning one's environment for potentially more urgent or stimulating tasks, rather than focusing on a single primary task (Stone 2009). Social media exacerbates this tendency by delivering a constant stream of distractions - new posts, reactions, messages, and other stimuli - making it difficult for users to sustain their attention on primary tasks (Norlock 2021). These platforms hijack users' attention with features such as the persistent notifications and the eye-catching red icons commonly used for alerts (Lewis 2017). Users are tempted to stay vigilant to these stimuli out of impulses (Lewis 2017). As this impulsive action is performed repeatedly, it will eventually become an unreflective habit. Among study participants, only one reported disabling TikTok notifications. The others receive regular prompts from the app about order updates, reactions to their contents, or incoming messages. Participants' responses to these notifications varied - some previewed and dismissed them, others tapped and opened the app to engage further. In all cases, these stimuli successfully diverted users' attention from their primary tasks, contributing to a pattern of fragmented focus. In this way, participants risk losing the ability to focus exclusively on their primary tasks at hand resulting from their continuous partial attention.

> **Q:** *How often does TikTok send you notifications?*
>
> **A:** *Quite often and almost every day. It will recommend information that I usually show interest in, or livestream sessions of shops, or new posts by the people I follow.*
>
> **Q:** *What do you do when you see those notifications?*
>
> **A:** *If I see something interesting, I will open TikTok. If it is not interesting for me, I just delete that notification from my phone screen.*
> — Interview 8

In addition to continuous partial attention, users' attention spans are also undermined by the brief duration of TikTok videos. These short-form contents rewire cognitive patterns, conditioning the brain to expect rapid, shallow input (Carr 2011). Over time, this erodes the ability to engage in deep, sustained thought or long-form reading - skills essential to intellectual development and knowledge acquisition (Carr 2011). In other words, one becomes accustomed to absorbing short and shallow contents if one constantly view such contents over a sufficient amount of time. In this way, as this process happens insidiously (Carr 2011), this is an unreflective habit. One participant observed that TikTok audiences prefer short videos and tend to disengage with longer ones. As users repeatedly consume such content without awareness or critical reflection, they lose the opportunity to build countermeasures that might mitigate cognitive decline. Since this erosion of focus is gradual and difficult to detect - and even harder to reverse (Carr 2011) - its long-term impact on both individual functioning and societal productivity is significant.

> **Q:** *How would you rate your overall experience with TikTok?*
>
> **A:** *I think it is quite neutral. Sometimes, I view inspiring videos, which I find okay. But people just use TikTok as a way to introduce their videos on other platforms, such as YouTube. This is because TikTok users are only interested in short videos. When I use TikTok, it is just like a way to escape my worries, but it doesn't make me feel happier.*
> — Interview 11

Third, unreflective habit formation can lead to identity rigidification (Marin 2025). This occurs when users become entrenched in certain beliefs, preferences, or values due to repetitive exposure to similar content (Marin 2025). TikTok's algorithm frequently recommends videos featuring women who are cute or conventionally attractive, which aligns with the aesthetic preferences of most participants. Although some participants admire idols for qualities like independence or altruism, they do not report performing any actions to replicate these behaviors. Their admiration stops at observation. However, many reported mimicking "cute" behaviors and using filters that enhance these traits in their own videos as they see videos showing cute and easy to replicate movements and expressions. Social approval in the form of likes and comments reinforces this behavior as they make participants feel recognized. As Attrill-Smith (2019) explains, extended role play in online environments can shape users' offline identities. While embodying a "cute" persona is not inherently negative, the concern is that TikTok's design encourages users to prioritize surface-level traits over more substantive values such as intelligence, autonomy, or career ambition - qualities that are arguably more beneficial for university students on the verge of adult working life.

In summary, participants' habitual TikTok use is triggered by emotional and situational cues and is reinforced through platform design and reward structures. As a result, users not only lose valuable time, but also suffer declines in attention and risk forming rigid identities that lack substantive values. Because these habits form without reflection, users are unlikely to recognize or mitigate these losses, leaving them increasingly vulnerable to the long-term consequences of unreflective habit formation.

6.2. TikTok: Hostile System

Interview findings indicate that TikTok may constitute a hostile system for formative agents by limiting their capacity for critical reflection. One clear consequence is what Marin (2025) terms unreflective habit formation, which expresses itself in two ways.

First, participants do not actively resist the formation of habitual TikTok use. Once the habit is established, it becomes an automatic behavior. TikTok has become their go-to app for relaxation moments. To most participants, the justification of boredom relief seems sufficient. As a result, few participants reflect deeply on their using TikTok as a habit.

Second, participants seldom questioned the implications of their usage. When TikTok becomes inaccessible, some report feelings of sadness or even aimlessness, as if something essential is missing. They often resumed usage at the earliest opportunity, either by viewing TikTok contents on their friends' device with them, or waiting for the app to function again. Only one participant admitted to bordering on addiction - but even she rationalized it by saying others around her behaved the same way. No participant appeared to challenge their emotional attachment or reliance on the platform.

Another consequence of TikTok's hostility is the lack of resistance participants pose to the platform's monetization of their attention. While participants mostly use TikTok for entertainment and relaxation, the app covertly holds captive of and monetizes their attention. Alarmingly, this exploitation is often perceived as beneficial. TikTok masks its extractive practices through three main mechanisms.

Firstly, TikTok makes beauty appear easily attainable. The abundance of beauty and cuteness that TikTok algorithm predominantly recommends on their personalized feed cultivates

desire in participants to emulate these appearances. This aspiration is then channeled into two actions - content creation and product consumption - both of which TikTok conveniently enables without users needing to leave the platform.

Secondly, TikTok leverages randomness to amplify the psychological rewards of social recognition. While there are tricks to improving a video's chances of going viral, there are no guarantees. This unpredictability drives participants to mimic trends with the hope of achieving short-lived stardom. The exhilarating sensation of becoming viral, despite its low chance, is irresistibly rewarding. So much so that participants are hooked: they stay tune for cutely choreographed trends, coordinate with their friends to re-create them, and hold their breath waiting for their posts to become viral.

One might argue that participants exercise autonomy in their methodical approach to content creation and purchasing, and that their actions satisfy criteria for autonomy as outlined in Section 4.3. Indeed, participants will not re-create videos of all the trends that they come across, nor do they buy everything they see on TikTok. Their autonomy is not fully undermined in this decision-making process. However, what is at stake is the trigger leading to this decision-making process: TikTok's content curation fuels desires that go largely unexamined. Rather than interrogating these manipulative triggers, participants appear to accept them because TikTok sugarcoats its tactics.

Thirdly, TikTok instills a new form of social pressure to reinforce participants' TikTok use through streaks. Breathing bonding values into streaks, TikTok creates a sense of connection among participants, compelling them to use TikTok daily to maintain their count of consecutive days of staying connected on TikTok. Most participants uncritically accept this assigned value without questioning whether there is any practical value brought about by these superficial virtual connections. They even go the extra miles to compel their friends to uphold streaks. They disregard the fact that with or without TikTok streaks, they are already friends in real life. In this case, participants showed little evidence in exercising their autonomy.

Under the lens of system hostility, the agent's autonomy is constrained because the system does not provide a fertile space for critical reflection. Instead, it narrows the agent's perceived set of choices. From the standpoint of strong substantive autonomy, the agent cannot be considered autonomous in such a restrictive and manipulative system.

7. STPA Analysis

7.1. Define Purpose of the Analysis

7.1.1. Challenges in Identifying Hazards

A fundamental aspect of STAMP is identifying system hazards - states of the system that may result in accidents and losses. STAMP operates on the principle that hazards must be identified and eliminated or controlled to ensure safety. Once hazards are established, the remaining components of the STPA analysis build on them. In STAMP frameworks, environmental conditions coupled with system hazards lead to accidents, which then result in losses. Hazards must describe system conditions that, in the worst-case environment, result in losses (Leveson and Thomas 2018).
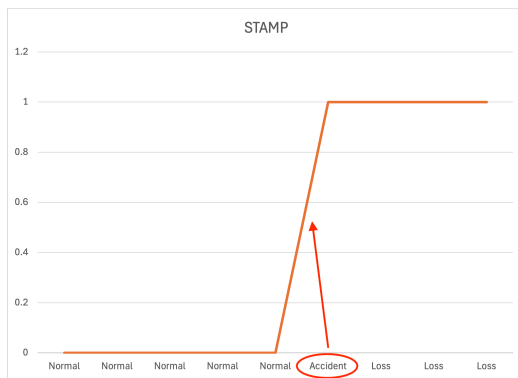
Figure 7.1.: Conventional Physical System Accidents and Losses
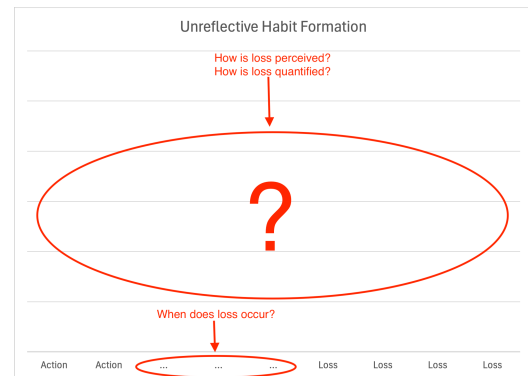
Figure 7.2.: Unreflective Habit Formation Harm and Resulting Losses Undefined Trajectory

Identifying hazards for unreflective habit formation is particularly difficult. The STAMP conceptual model is not easily applicable to losses that are imperceptible, unquantifiable, and gradually accumulate over time through habitual actions. For example, users can measure their screen time, but cannot easily perceive or quantify the erosion of their focus or the merging of their online and offline personas. One or two sessions may not produce harm. But after the nth time, the losses might begin to take effect - subtly, perhaps even dismissibly - and they continue to accumulate in the following sessions. At some y^{th} instance, the damage may become too severe to simply undo. But when exactly is that n^{th} or y^{th} time? Unlike traditional accidents with clear events, such as explosions in aerospace, these losses develop subtly and without a clear threshold, making it difficult to determine hazardous states.

To apply STPA, there must be a way to measure losses so that a threshold at that y^{th} instance can be identified. Given the limited capacity of the researcher and the scope of this study,

the task of defining and justifying such a threshold is deferred to experts in their respective domains. The threshold serves two important purposes in hazard identification: (1) it marks a tipping point that allows the corresponding system state or conditions to be identified, and (2) it creates a pseudo-accident condition, in which the loss is considered to have "occurred" once the threshold is crossed.

Another approach to applying the STAMP model is to center harm conceptualization on unreflective habit formation itself rather than on the resulting losses. In this framing, the process of unreflective habit formation is treated as an accumulation of recurrent "accidents", where each accident is equivalent to a single action enactment. The more frequently an action is performed within a stable context, the more likely it is to be repeated in the future, eventually becoming a habit (Pinder et al. 2018). Each individual incident contributes to exacerbating the resulting losses. Therefore, preventing each unreflective action enactment directly supports the prevention of unreflective habit formation, thereby mitigating resulting losses. In Figure 7.3, which illustrates the problem conceptualization in Section 7.1.4, each of these accident, where contexts leads to impulsive TikTok use under the heavy influence of TikTok, is referred to as "individual incident".

Given that individual actions are more tangible and observable than abstract, pernicious losses, this approach is more manageable than centering the conceptualization on the losses themselves. Consequently, the research adopts this approach in its application of the STAMP conceptualization and STPA method - focusing on the prevention of each instance in which users engage with TikTok under their regular usage contexts.

7.1.2. System Boundary

For the harm of unreflective habit formation, the system boundary in the STPA analysis is defined to encompass two primary components: users and **TikTok**. This framing highlights the dynamic interaction between the human elements (users) and the sociotechnical system (TikTok), emphasizing how platform design and user behavior co-produce the conditions that may lead to unreflective habit formation. The system is illustrated in the functional control diagram in Figure 7.5 in Section 7.2.2.

7.1.3. Losses, Hazards, and System-Level Constraints

The three losses, as described in Section 6.1, are rephrased here for conciseness:

- Users lose time as they habitually use TikTok. **[L1]**

- Users lose the capacity to focus as they habitually use TikTok. **[L2]**

- Users face identity rigidification as they habitually use TikTok. **[L3]**

According to Leveson and Thomas (2018), hazards must be expressed as system states or conditions to be avoided, as they can lead to losses under worst-case environment. Based on the findings presented in Chapter 5, the process through which users incur the three losses as a result of their habitual TikTok use can be described as follows:

- Users experience boredom, mental fatigue, stress, or simply have free time. In these states, they default to their habitual TikTok use to unwind. The platform rewards them with entertainment and relaxation, reinforcing the habit and contributing to worsening the losses.

In this context, the system hazardous state is characterized by users experiencing boredom, fatigue, stress, or spare time and turning to TikTok out of habit. Focusing on the individual action enactment as explained in Section 7.1.1, the hazard is therefore phrased as:

- Users receive stimuli, such as boredom, mental fatigue, stress or having free time, and they unreflectively use TikTok in response. **[H]**

Although this hazardous state centers on the users component, hastily assigning blame solely to users would be misguided. Hazards are emergent properties of a system, arising from complex interactions among its components (Leveson 2012). In the case of unreflective habit formation on TikTok, the platform is deliberately engineered to maximize attention capture and habit reinforcement (Mik 2016; Lewis 2017; Susser, Roessler, and Nissenbaum 2019; Sahebi and Formosa 2022; Esposito and Ferreira 2024). TikTok must therefore bear an equal - if not greater - share of accountability for generating these harms and addressing them.

Focusing on holding TikTok accountable, the system-level constraint derived from the identified hazard is defined in this analysis as:

- The system must not encourage users to unreflectively use TikTok in response to stimuli such as boredom, mental fatigue, stress or having spare time. **[S]**

7.1.4. Problem Conceptualization

Using the mechanism of unreflective habit formation on social media identified in Section 4.4, the problem of unreflective habit formation can be framed in STAMP terminology at a high level as follows:

> Contexts cause triggers for impulsive TikTok use. Undermining user autonomy, TikTok intensifies the trigger **[H]**. As users repeatedly respond to contexts with impulsive TikTok use, their impulsive TikTok use becomes an unreflective habit. This unreflective habit formation results in three losses, namely Time **[L1]**, Focus capability **[L2]**, and Identity rigidification **[L3]**.

Figure 7.3 visually illustrates this problem conceptualization. As previously discussed in Section 7.1.1, the component "Repeated individual incidents" is central to both the conceptualization using the STAMP model and the analysis using the STPA method.
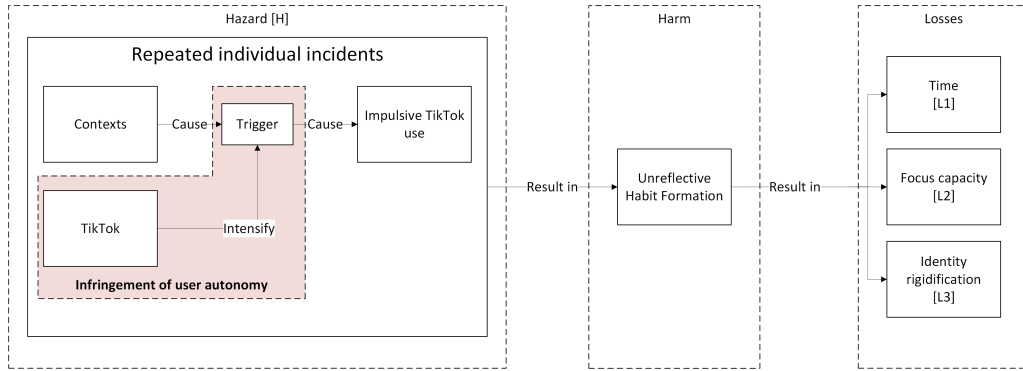
Figure 7.3.: Problem conceptualization: TikTok infringes user autonomy by intensifying the trigger leading to impulsive TikTok use. Repeated impulsive TikTok usage under specific contexts results in unreflective habit formation. Unreflective habit formation results in losses.

7.2. Model the Control Structure

7.2.1. Hierarchical Control Structure

In Vietnam, system safety for social media platforms like TikTok is jointly enforced by the Government and by the platforms themselves, with TikTok T&S operations serving this role in the present case. The hierarchical control structure that shows system safety mechanisms by the Government and TikTok is shown in Figure 7.4.
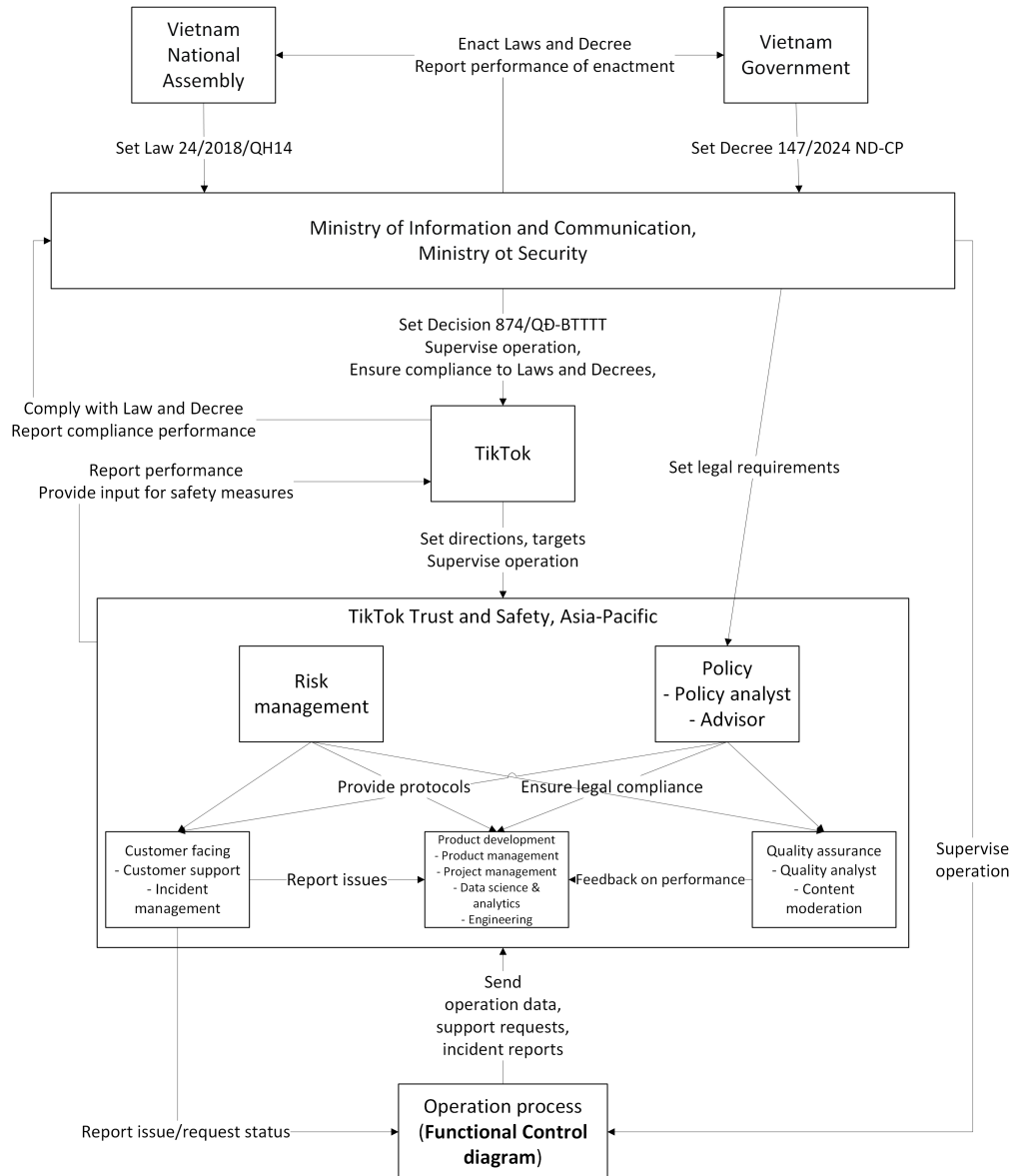
Figure 7.4.: System Control Structure by Vietnam Government and TikTok T&S

Government Oversight

The Vietnamese Government and National Assembly maintain cybersecurity and online safety through three main legal frameworks:

- **Law 24/2018/QH14**: The Cybersecurity Law, announced by the Vietnam National Assembly in 2018 and effective from 2019, regulates activities for protecting national security and social order in cyberspace. A major focus is on content moderation and the removal of harmful or illegal content (Thu Vien Phap Luat 2025a).

- **Decree 147/2024/NĐ-CP**: Announced by the Vietnam Government in 2024, this Decree prescribes measures for the management, provision, and use of Internet services and online information. It outlines detailed approval procedures for social media providers, mandates traffic statistics reporting, and requires cooperation in moderating illegal content (Thu Vien Phap Luat 2025b).

- **Decision 874/QĐ-BTTTT**: Issued by the MIC in 2021, this Code of Conduct on Social Media explicitly requires service providers to prevent exploitation and emotional or psychological abuse of users (Thu Vien Phap Luat 2025c).

Overall, the MIC is tasked with managing cyberspace, including TikTok's operations in Vietnam, often in collaboration with other ministries (Ministry of Science and Technology 2023; Bao Chinh Phu 2023). Following investigations, the MIC demands TikTok to rectify malpractices (Bao Chinh Phu 2023).

These three legal frameworks are not yet explicit regarding user exploitation or phenomena like unreflective habit formation, a gap that persists globally. Nevertheless, recognizing that children are particularly vulnerable, Vietnam is increasingly imposing explicit paternalistic policies toward minors, such as mandatory age verification, account deletion for users under 13, usage time restrictions, and prohibition of monetization for minors (Thu Hang 2023). For adults, direct paternalistic interventions are more controversial, as such policies could be seen as authoritarian and limiting of freedom. This calls for more scientific research on the pernicious harm of unreflective habit formation for the government to derive appropriate policies for the wider population, not just children.

TikTok T&S Team

User safety at TikTok is overseen by the TikTok T&S team (Keenan 2021), with regional oversight in Vietnam handled by the Asia-Pacific (APAC) branch (TikTok 2025t).

TikTok (2025l) lists the open positions at TikTok, and filtering the listings to roles within the TikTok T&S team in APAC region reveals a broad array of specialized teams and roles. The job postings retrieved from the TikTok (2025l) reveals that the TikTok T&S organization consists of multiple specialized teams with distinct responsibilities that work closely together across different functions.

- Product development: Includes product managers, project managers, engineers (including machine learning engineers), and data scientists. They build and refine safety-related features, internal tools, and AI moderation models (TikTok 2025b; TikTok 2025q; TikTok 2025p; TikTok 2025r; TikTok 2025s; TikTok 2025e; TikTok 2025u; TikTok 2025o; TikTok 2025d).

- Quality assurance: Quality analysts and content moderators verify and refine the outputs of automated moderation systems (TikTok 2021; TikTok 2025c; TikTok 2023; TikTok 2025g; TikTok 2025m).

- Legal and policy compliance: Legal advisors and policy analysts ensure TikTok aligns with local regulations, especially regarding sensitive issues like child safety (TikTok 2025h; TikTok 2025n).

- Risk management: Focused on predicting and preventing risks (TikTok 2025f; TikTok 2025t).

- Customer support: Customer service agents and incident managers handle user complaints and escalate issues when necessary (TikTok 2025j; TikTok 2025k).

While TikTok demonstrates a commitment to safety through organizational specialization, evaluating the effectiveness of these efforts requires further investigation. Importantly, Tik-Tok's safety focus revolves heavily around content moderation and fraud prevention. Issues like user habit formation or excessive platform dependence receive little explicit attention. This is mainly due to the fact that the platform's business incentives favor prolonged user engagement, suggesting that addressing phenomena like unreflective habit formation runs counter to TikTok's financial interests.

7.2.2. Functional Control Diagram

The functional control diagram is a visual model that illustrates both the controlled process and its associated controllers. The functional control diagram for unreflective habit formation on TikTok is constructed based on two sources: (1) the literature review on how habits are formed, as discussed in Section 4.4, and (2) desk research into relevant TikTok job listings, particularly those related to the TikTok T&S team. Figure 7.5 presents the resulting control diagram, and Figure 7.6 illustrates the controlled process.
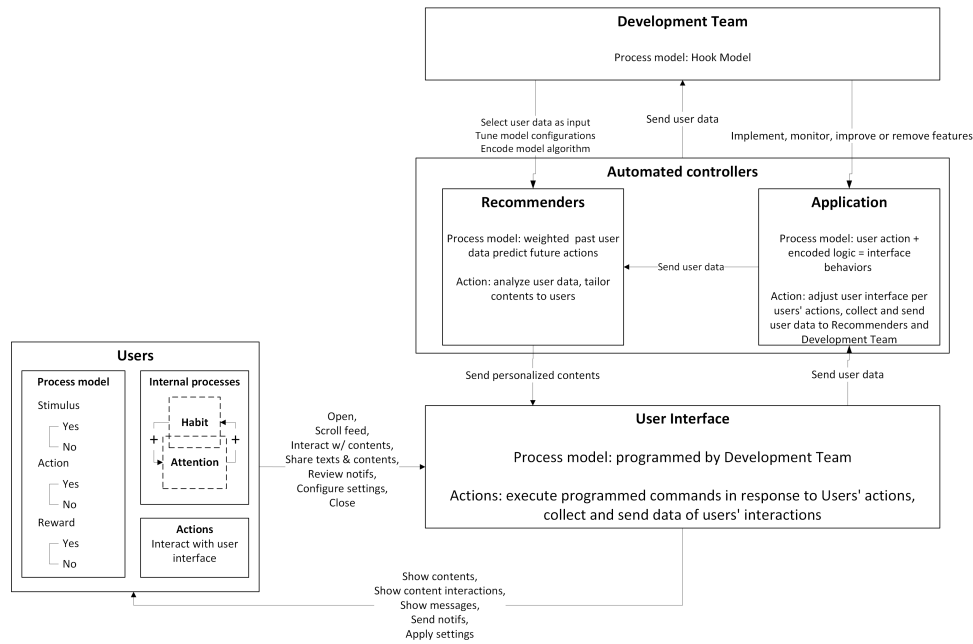
Figure 7.5.: Functional Control Diagram for Unreflective Habit Formation on TikTok

In the control diagram, regarding controllers, the components are as follows:

- UI: This encompasses the features presented to users and the content displayed to them, via either the web or mobile app. It is the primary touchpoint for user interaction and the space where the product development team shapes what users can do on TikTok and how they experience the platform.

- **Automated controllers**: The system includes two automated controllers. First, the recommender algorithms (labeled as **recommenders** in Figure 7.5) mediate content flow between users and the broader TikTok community. Second, the TikTok application (labeled as **application** in Figure 7.5) governs the operation of the UI - for example, when a user taps the "heart" icon on a post, the application automatically increments the like counter and displays visual confirmation of the action. The product development team control the UI through the application.

- **TikTok product development team**: Represented as the **development team** in Figure 7.5, this human controller manages both automated controllers. The team is responsible for configuring recommenders to enhance the accuracy of the personalized contents shown to users. The team control the application by collecting user data to discover, design, implement, monitor, improve, or remove features. Their decision-making is driven by engagement metrics (Haidt 2023), and their process model follows the Hook Model, which was introduced in the Section 4.4 and illustrated in Figure 4.2.

- **Users**: Users function as human controllers by taking actions within the UI. These actions are designed by the product development team. Assigning a traditional process model to users is complex, as it typically reflects a controller's mental model of the system. In the context of social media, user behavior is shaped by the reward cycle - though users are often unaware of this mechanism. Therefore, the user's process model is a modified version of the reward cycle: upon encountering internal or external stimuli prompting social media use, users perform actions on the UI, which in turn provides them with a mental or emotional reward. Regarding the **internal processes** component inside the users, there are two internal mental processes that all of the controllers, to varying extents, control, or at least strive to. These processes are **habit** and **attention**, which is further explained below and illustrated in Figure 7.6.
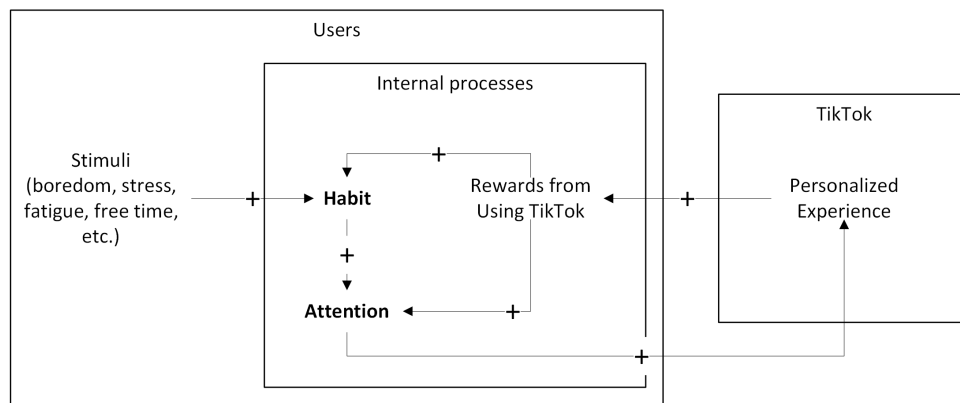
Figure 7.6.: Users's **Internal processes**: Reinforcing loop of **Habit** and **Attention**

Regarding controlled processes, there are two processes in question, both of which are users' internal mental processes. The first process is users' unreflective habitual TikTok use (referred to as **habit** in Figure 7.6), which is a result of unreflective habit formation as explained in Section 4.4 and Section 6.1. The second process is users' highly engrossed engagement with TikTok (referred to as **attention** in Figure 7.6), the mechanism of which is explained in Section 4.4. These two internal processes are closely intertwined in two main ways.

Firstly, they are important in driving the direction of the products. TikTok's internal product development decisions are driven primarily by user engagement metrics (Haidt 2023), among which are measures of the number of active users and the duration of their sessions. The more users compulsively open TikTok, the higher the number of active users - resulting in a larger audience pool. The more engaged users are with TikTok, the longer they stay, which increases the volume of content exposure. The better these metrics perform, the more valuable TikTok becomes to advertisers, and thus the greater the platform's monetary gains. It is essential for TikTok to maintain strong performance across both metrics, as it indicates not only a large user base, but also successful competition against alternative platforms (Haidt 2023).

Secondly, together they form a closed causal loop in which each reinforces the other, which is illustrated in Figure 7.6. Users compulsively open TikTok out of habit, intending only to spend a few moments, but end up staying for extended periods and paying full attention to TikTok (Haidt 2023). The longer users remain engaged, the more data TikTok collects and uses to personalize user experiences (Haidt 2023), making the pseudo-rewards more compelling. Users are enticed by the neurological reward cycle to anticipate and seek further rewards, thereby reinforcing the habitual action that yields those anticipated rewards (Esposito and Ferreira 2024). In short, users develop the unreflective habit of using TikTok in response to stimuli because it provides rewards. When users open TikTok, they remain engaged for an uncontrollable amount of time. Their extended engagement enables the system to tailor even better rewards, ensuring users remain engaged and return again - thus strengthening the habit.

The functional control diagram in Figure 7.5 reveals the intricate interplay of different system components, including controllers and controlled processes. While the development team architects the recommenders, application, and UI, the users' interaction with the system, in reality, shapes these design choices. Although the recommenders and the application control the UI, they do so in response to the users' inputs. The development team, with their decision-making powers and acting through automated controllers, proactively governs the system by designing and implementing its boundaries and behaviors. The users, in contrast, play the role of passive controllers, acting only within the constraints set by the system's architecture.

7.3. Identify UCA

7.3.1. UCA

Each UCA specifies the controller, the type of unsafe condition, the control action and the context (Leveson 2012). There are four types of unsafe condition when UCA happens, namely: 1) not providing the control action causes hazard; 2) proving the control action causes hazards; 3) providing the control action too early, too late or out of order causes hazards; 4) stopping control actions too soon or applying them too long causes hazards (Leveson 2012). Each of these conditions is listed as a column entry in the Table 7.1. Since there are no actions falling under the condition "too early, too late, or out of order", this column is omitted from the Table 7.1. Based on the functional control diagram (Figure 7.5) and interview data, the following UCA have been identified and are summarized in Table 7.1.

In Table 7.1, while some UCA lead directly to losses, others serve to reinforce the hazard **[H]**. As **[H]** recurs over time, it exacerbates the resulting losses, thereby establishing an indirect link between these UCA and the losses.

In airplanes, components such as wheel brakes and wing spoilers are designed with the primary goal of ensuring safety. By contrast, as evidenced by the UCA entries in Table 7.1, most features on TikTok are designed to maximize user engagement rather than promote user safety. In the context of unreflective habit formation, these features can therefore be considered inherently hazardous.

Table 7.1.: UCA relating to the Controlled Processes of **Users' Habits** and **Attention** on TikTok

Controller	Action	Not providing causes hazard	Providing causes hazard	Stopped too soon, applied too long
Application	Send notifications about TikTok activities (post reactions, order status, messages from friends, new contents, etc.)		UCA 1: Application sends notifications about TikTok activities when users use TikTok. **[H]** UCA 2: Application sends notifications about TikTok activities when users don't use TikTok. **[H]**	UCA 3: Application keeps notifications on the device screen until users interact with them. **[H]**
Application	Show contents			UCA 4: Application shows users continuous contents when their TikTok session has already been too long. **[L1]**
Application	Show content interactions		UCA 5: Application shows users the interactions to their posts when users publish their posts. **[H]**	
Application	Show messages from friends	UCA 6: Application does not allow users to resume their message streaks when users forget to message their friends one or a few days. **[H]**	UCA 7: Application shows users messages from friends to remind them to maintain messaging streaks. **[H]**	

Controller	Action	Not providing causes hazard	Providing causes hazard	Stopped too soon, applied too long
Application	Apply users' settings	UCA 8: Application allows users to bypass screen time management alarms and to resume using TikTok when these alarms go off. **[H, L1]**		
Application, Software	Send user data		UCA 9: Application collects and sends all user data to automated controllers and development team without explicitly informing users what and how it will be used. **[H, L3]**	UCA 10: Application doesn't stop collecting and sending user data to automated controllers and development team, even when there is already sufficient data to recommend contents to users. **[H, L3]**
Recommenders	Send personalized contents		UCA 11: Recommenders keep showing users contents aligning with their preferences when screen time management alarms go off. **[H, L1, L3]** UCA 12: Recommenders favor short over extended contents. **[H, L2]**	UCA 13: Recommenders keep showing users contents aligning with their preferences when their TikTok session has already been too long. **[H, L1, L3]**

Controller	Action	Not providing causes hazard	Providing causes hazard	Stopped too soon, applied too long
Users	Configure screen time management features, such as screen time limit, screen time break, sleep reminder	UCA 14: Application doesn't show or introduce to users screen time management settings (screen time limit, screen time break, sleep reminders) when users are unaware of and do not intentionally look for these settings. **[H, L1]**		
Users	Configure collected data and its usage	UCA 15: Application doesn't allow users to specify the type of data to be collected and used to enhance their experience when they are unaware of the data collection mechanism. **[L3]**		
Development team	Select user data as input to recommenders		UCA 16: Development team relies on engagement metrics to decide the type of data and their weight as inputs when explaining their decisions to leaders. **[H, L1, L2, L3]**	
Development team	Implement, monitor, improve or remove features		UCA 17: Development team relies on engagement metrics to shape the application when explaining their decisions to leaders. **[H, L1, L2, L3]**	

7.3.2. Controller Constraints

Controller constraints are behaviors of controllers designed to prevent UCA from occurring (Leveson and Thomas 2018). Based on the analysis in Table 7.1, the corresponding controller constraints have been identified and are presented in Table 7.2.

Table 7.2.: UCA and Corresponding Controller Constraints

UCA	Controller Constraint
UCA 1: Application sends notifications about TikTok activities when users use TikTok. **[H]**	C-1: Application must avoid sending non-essential notifications when users use TikTok. **[H]**
UCA 2: Application sends notifications about TikTok activities when users don't use TikTok. **[H]**	C-2: Application must avoid sending non-essential notifications when users don't use TikTok. **[H]**
UCA 3: Application keeps notifications on the device screen until users interact with them. **[H]**	C-3: Application must auto-dismiss or archive non-essential notifications after a reasonable period. **[H]**
UCA 4: Application shows users continuous contents when their TikTok session has already been too long. **[L1]**	C-4: Application must introduce stopping cues or session limits after a long session. **[L1]**
UCA 5: Application shows users the interactions to their posts when users publish their posts. **[H]**	C-5: Application must not overemphasize social validations when users publish their posts. **[H]**
UCA 6: Application does not allow users to resume their message streaks when users forget to message their friends one or a few days. **[H]**	C-6: Application must offer streak recovery to reduce pressure from maintaining it. **[H]**
UCA 7: Application shows users cues to remind them to maintain messaging streaks when they use TikTok. **[H]**	C-7: Application must not nudge users to maintain message streaks daily. **[H]**
UCA 8: Application allows users to bypass screen time management alarms and to resume using TikTok when these alarms go off. **[H, L1]**	C-8: Application must not allow users to bypass screen time limit, screen time break and sleep reminder alarms when they go off. **[H, L1]**
UCA 9: Application collects and sends all user data to automated controllers and development team without explicitly informing users what and how it will be used. **[H, L3]**	C-9: Application must disclose to and seek consent from users for all data sent to automated controllers and development team. **[H, L3]**
UCA 10: Application doesn't stop collecting and sending user data to automated controllers and development team, even when there is already sufficient data to recommend contents to users. **[H, L3]**	C-10: Application must not continue excessive data collection once there is sufficient data to curtail contents to users. **[H, L3]**
UCA 11: Recommenders keep showing users contents aligning with their preferences when screen time management alarms go off. **[H, L1, L3]**	C-11: Recommenders must reduce contents aligning with users' preferences when the user-specified screen time limit or sleeping time is approaching, and the alarms for these settings go off. **[H, L1, L3]**
UCA 12: Recommenders favor short over extended contents **[H, L2]**	C-12: Recommenders must not favor brief over extended contents. **[H, L2]**

UCA	Controller Constraint
UCA 13: Recommenders keep showing users contents aligning with their preferences when their TikTok session has already been too long. **[H, L1, L3]**	C-13: Recommenders must reduce contents aligning with users' preferences when their Tik-Tok session has already been too long. **[H, L1, L3]**
UCA 14: Application doesn't show or introduce to users screen time management settings when they are unaware of these settings. **[H, L1]**	C-14: Application must introduce screen time management tools to users when users are unaware of and do not intentionally look for these settings. **[H, L1]**
UCA 15: Application doesn't allow users to specify the type of data to be collected and used to enhance their experience when they are unaware of the data collection mechanism. **[L3]**	C-15: Application must allow users to specify the type of data to be collected and used to enhance their experience. **[L3]**
UCA 16: Development team relies on engagement metrics to decide the type of data and their weight as inputs when explaining their decision to leaders. **[H, L1, L2, L3]**	C-16: Development team must not rely solely on engagement metrics to decide the type of data and their weight as inputs when explaining their decision to leaders. **[H, L1, L2, L3]**
UCA 17: Development team relies on engagement metrics to shape the Application when explaining their decision to leaders. **[H, L1, L2, L3]**	C-17: Development team must not rely solely on engagement metrics to shape the Application when explaining their decision to leaders. **[H, L1, L2, L3]**

7.4. Identify Loss Scenarios

"A loss scenario describes the causal factors that can lead to the unsafe control actions and to hazards" (Leveson and Thomas 2018, p.42). As outlined by Leveson and Thomas (2018), loss scenarios generally belong to four categories, namely:

- Unsafe controller behavior

- Causes of inadequate feedback/information

- Control path

- Other factors related to the controlled process

The categories of controlled path, feedback/information, and controlled processes primarily concern the malfunctioning of physical system components. However, speculating loss scenarios from physical failure is not productive in the case of unreflective habit formation, for two main reasons. First, the physical components and underlying algorithms - including the recommenders, application, and UI - are thoroughly designed, implemented, tested, and monitored by the development team to minimize bugs and optimize performance. These components are not accidental flaws but deliberate design choices, skillfully architected to reinforce automated habits and user engagement (Esposito and Ferreira 2024). In this context, the UCA are intentional and informed decisions, not system errors - they are developed and deployed in alignment with engagement-driven objectives. Second, any issue with physical components is likely to be detected rapidly through performance monitoring and resolved quickly to maintain a seamless user experience - an essential factor in sustaining users' attention and engagement (Haidt 2023).

Because the UCA identified in Section 7.3.1 result from deliberate choices by human controllers, the category of unsafe controller behaviors is the most relevant for explaining the unsafe nature of the UCA listed in Section 7.3.1. Therefore, this research focuses on that category to identify associated loss scenarios.

This research focuses on UCA in three key areas to identify loss scenarios. First, given TikTok's emphasis on user engagement metrics, loss scenarios will be identified for UCA 16 and UCA 17. Second, given the underutilization of safety measures reported in the interview findings in Section 5.2 - despite their potential to help users regulate their TikTok usage - loss scenarios will be identified for UCA 8 and UCA 14. Third, based on users' responses to streaks discussed in the interview results in Section 5.6, the research will identify loss scenarios for UCA 6 and UCA 7.

7.4.1. Loss Scenarios 1

UCA 1 to UCA 15 outline features controlled by the automated controllers - namely, application and recommenders - and presented to users via the UI. These TikTok features are thoroughly designed, implemented, tested, and optimized to maximize user engagement metrics. They are intentional design choices architected to encourage repeated use and prolonged engagement with the platform.

In the context of unreflective habit formation, these features are inherently hazardous because they reinforce the action of using TikTok under stable contexts - such as when users experience mental fatigue or seek entertainment and relaxation. The functional control diagram in Figure 7.5 shows that the development team is responsible for managing the behavior of the automated controllers and the appearance of the UI. Put differently, UCA 16 and UCA 17 are the root causes of the preceding UCA 1 to UCA 15.

However, it would be reductive to place sole responsibility on the development team. Their reliance on engagement metrics reflects the priorities set by their leadership (Haidt 2023). As shown in the hierarchical control structure in Figure 7.4, the development team operates at the execution level and reports to senior TikTok leadership, who emphasize engagement above other considerations. User engagement metrics is emphasized because TikTok offers its service to users for free and monetizes by advertisement and e-commerce. The more frequently and longer users use its platform, the more profit TikTok earns. The loss scenario is then summarized as follows:

> **Loss Scenario 1 for UCA-16 and UCA-17:** Users open TikTok. The UI and automated controllers are encoded to compel users to use TikTok compulsively and stay highly engaged for extended periods [**UCA-1** to **UCA-15**]. As a result, user engagement metrics perform well. The development team relies on these metrics when building the UI and automated controllers and when explaining decisions to leadership [**UCA-16**, **UCA-17**]. This is because TikTok leadership places high emphasis on these metrics. As TikTok monetizes on advertisement and e-commerce, user engagement directly affects its profit. **[H, L1, L2, L3]**

7.4.2. Loss Scenarios 2, 3

TikTok has implemented safety features intended to help users manage their screen time and usage. While TikTok is fully aware that these screen time management features are largely

ineffective, they exist primarily to demonstrate to policymakers TikTok's effort to ensure user safety (Haidt 2023). These features are only half-heartedly enforced because serious enforcement would negatively impact user engagement metrics.

There are two primary reasons why these features are underutilized. First, users are often unaware of their existence, as the features are discreetly buried within the settings menu [UCA 14]. Second, their implementation is largely symbolic: even when users are aware of these tools, they can easily bypass them and continue their TikTok sessions [UCA 8]. Using a combination of features, TikTok makes it difficult for users to exit the app once prompted by these reminders. For instance:

- The application presents continuous content in the UI [UCA 4] that strongly aligns with users' interests based on the recommenders' suggestion [UCA 11], [UCA 13].

- The application displays notifications about post interactions [UCA 5], prompting users to engage further.

- The application shows notifications in the UI, demanding users' attention [UCA 1], [UCA 2], [UCA 3].

The resulting loss scenarios are summarized as follows:

Loss scenario 2 for UCA 14: TikTok implements safety features such as screen time limits, screen time breaks, and sleep reminders. However, the application does not proactively introduce these settings to users who are unaware of them [UCA 14]. Despite recognizing the low adoption of these features, TikTok has not improved their effectiveness, prioritizing engagement metrics instead. So, users spend an excessive amount of time using TikTok. **[H, L1]**

Loss scenario 3 for UCA 8: While safety features exist, users can easily bypass them and resume their TikTok sessions. Given the abundance of features that promote impulsive and continuous use [UCA 1, UCA 2, UCA 3, UCA 4, UCA 5, UCA 11, UCA 13], users become deeply absorbed in their sessions, making it difficult to moderate their usage. Despite being aware of the ineffectiveness of these tools in helping users regulate their screen time, TikTok has made no significant effort to enhance them, again due to concerns over compromising engagement metrics. So, users spend an excessive amount of time using TikTok. **[H, L1]**

7.4.3. Loss Scenarios 4

The streak feature introduces a new form of pressure that encourages users to engage with TikTok daily. The application displays visual indicators reminding users to maintain messaging streaks with their friends [UCA 7]. By assigning value to these streaks, users feel compelled to preserve these perceived friendship milestones - and, in turn, often nudge their friends to reciprocate. To maintain a streak, users must log in and interact with TikTok every day [UCA 6]. In this way, the streak mechanism reinforces TikTok as a daily habit. This behavior contributes directly to the platform's user engagement metrics, such as daily and monthly active users.

TikTok did not originate the streak concept. Long before its adoption by TikTok, Snapchat had implemented streaks to reward ongoing user interactions. Duolingo also employs the

feature to motivate daily lesson completion. Streaks therefore exemplify how companies adopt and repurpose successful user retention strategies from other platforms.

Loss scenario 4 for [UCA 6] and [UCA 7]: TikTok adopts proven user retention strategies from other platforms, such as streaks. The application displays visual streak indicators to users, encouraging them to maintain their streaks [UCA 7]. In response, users must engage with TikTok daily to preserve these streaks [UCA 6]. This mechanism reinforces reinforce using TikTok as a daily habit [H].

8. Policy Recommendations

The harm of unreflective habit formation is eminent for TikTok users, including Vietnamese university students. This harm is an expression of the loss of user autonomy by social media in general. In addition to losing autonomy, unreflective habit formation also results in loss of time, ability to focus and meaningful identity transformation. These abstract losses have practical implications to not only individuals but also society as a whole. One doesn't have to look further than the cases where social media is used to manipulate people's opinions and emotions to interfere with democratic processes (Susser, Roessler, and Nissenbaum 2019).

8.1. Actor Roles and Responsibilities

The two most powerful actors in the socio-technical TikTok system in Vietnam are the Vietnam Government (herein referred to as the Government) and TikTok. The former holds the power of regulation, and the latter implementation. Among these two actors, the responsibility leans heavily on the Government for two main reasons. Firstly, TikTok cannot be counted on to implement safety measures to prevent users from unreflective habit formation as it goes against their business model. The more severe and widespread unreflective habit formation happens among TikTok users, the higher the user engagement metrics, and the more lucrative TikTok becomes from advertisement and consumerism. TikTok is fully aware of problematic usage from its users, such as compulsive use, deteriorating mental well-being, or exposure to harmful contents, etc., yet they not only turn a blind eye on remedies but also strive to monetize users even more (Haidt 2023). Secondly, the Government shows intentions to keep TikTok under tight control and does take measures to do so (Ministry of Science and Technology 2023; Ministry of Information and Communications 2023). In response to the scrutiny from the Government, TikTok shows willingness to comply (Thu Hang 2023). When clearer regulations are imposed, TikTok T&S Policy team will be responsible for enforcing compliance throughout the organization.

8.2. Recommendations to the Government

Due to the exploratory nature of this research, only a preliminary list of recommendations can be offered to support the Government in initiating mitigation strategies for unreflective habit formation on platforms such as TikTok. The research will utilize the output of STPA, namely system-level constraints, UCA and loss scenarios, to propose a set of recommendations for the Government.

8.2.1. Recommendation 1: Leverage Science for Contextualization of Harms

It is crucial to contextualize the harms of unreflective habit formation in the Vietnamese context to ensure that future interventions are culturally and socially appropriate. To begin, the Government can leverage scientific research that explores user behavior patterns on social media at scale. This can help illuminate specific contexts that foster, and consequences that result from, unreflective habit formation in Vietnam. Another urgent area involves the mechanisms through which social media platforms hijack users' attention and intention. Given the rapid evolution of social media, research into these mechanisms must be ongoing to maintain a current understanding of the risks involved.

8.2.2. Recommendation 2: Recognize Unreflective Habit Formation on Legislative Documents

The Government should recognize unreflective habit formation, along with the loss of user autonomy, as serious harms requiring mitigation or prevention in legislative documents, namely Decree 147/2024/ND-CP and Decision 874/QD-BTTTT. Currently, content moderation - explicitly governed by official legislative and regulatory documents - is the primary focus and measure pursued by the Government. While content moderation plays a key role in maintaining social order, it is not sufficient to address all forms of problematic social media usage, one of which is unreflective habit formation.

To do this, the Government can consult the system-level constraint identified in Section 7.1.3, which is as follows: **The system must not encourage users to unreflectively use TikTok in response to stimuli such as boredom, mental fatigue, stress, or having spare time. [S]**. While this system-level constraint is not sufficiently refined for direct adoption, it may serve as a conceptual foundation for drafting tailored policy clauses that align with governmental priorities and regulatory frameworks. The loss scenarios in Section 7.4 provide further insights to improve the specificity and practicality of these policies. The recommendations based on specific loss scenarios are continued below.

8.2.3. Recommendation 3: Utilize User Wellness Metrics

The root cause of loss scenario 1 in Section 7.4.1 is the overemphasis on user engagement metrics. To counter this, the Government can explore the feasibility of mandating user wellness metrics. Two main challenges arise in this approach. First, user wellness is not directly measurable via TikTok. However, the Government can consult experts to identify viable proxies. For example, to address poor sleep quality caused by late-night TikTok usage, a reduced count of active users during nighttime hours - or improved compliance with sleep reminders - could serve as indicators. Similarly, reduced frequency and duration of daily active sessions could signal mitigated impulsive use. Regardless of the proxy selected, decisions about measurement must be grounded in scientific research contextualized for Vietnam.

The second challenge concerns access to internal knowledge on user wellness. TikTok's internal research teams are aware of its psychological effects (Haidt 2023), but this information

is seldom made public. If the Government employs legislative tools to compel TikTok to disclose such findings, the platform's policy team may respond selectively. Therefore, independent observers - such as NGOs or research institutions - should be empowered to investigate these impacts. Once a framework for user wellness metrics is defined and implemented, the Government can require platforms like TikTok to adjust their practices accordingly.

8.2.4. Recommendation 4: Oblige Strict Enforcement of Screen Time Management Features. Raise User Feature Awareness Via Education

Loss scenarios 2 and 3 in Section 7.4.2 highlight TikTok's symbolic implementation of screen time management features. The Government should compel TikTok to take these safety features seriously. Regarding loss scenario 2, TikTok must be required to ensure that these features are properly introduced and easily accessible. This may include reducing the steps needed to access these features rather than burying them deep within settings. At the same time, the Government should educate users about the importance of these tools to increase their adoption and proper use.

With respect to loss scenario 3, TikTok must be obligated to enforce these features rather than allow easy circumvention. When users set sleep or screen time limits, they are essentially committing to responsible platform use. TikTok, by allowing these settings to be bypassed - while maintaining a highly engaging and attention-capturing interface - undermines that commitment and user autonomy. On this basis, the Government should prohibit such bypass mechanisms.

8.2.5. Recommendation 5: Enforce Strict Regulations on Features Promoting Daily Engagement. Raise User Awareness Via Education

The streak feature is one among many on TikTok that compel users to engage with the platform compulsively and regularly. As shown in loss scenario 4 in Section 7.4.3, streak - like many other TikTok features - is not an original invention, but rather a concept replicated from other platforms. However, banning streak is not a straightforward solution. Duolingo uses streaks to promote consistent learning, and fitness apps implement them to encourage regular exercise. Although banning streaks may create obstacles for TikTok, it would not prevent the platform from introducing alternative features that similarly require daily user engagement, using justifications similar to Duolingo or fitness apps. Therefore, it is essential that the Government both strictly regulate features that demand daily engagement on TikTok and educate users about their potentially harmful nature. Raising awareness of the exploitative and manipulative design of such features can help dissuade at least a portion of the population from adopting or endorsing them uncritically.

8.2.6. General Recommendations

Beyond these specific recommendations linked to the four loss scenarios, the list of UCAs and controller constraints in Section 7.3.2 can support the Government in identifying other harmful system behaviors and formulating corresponding constraints. These insights may

be used as a starting point for future loss scenario identification and the drafting of mitigating regulatory measures.

Importantly, these measures should not be limited to TikTok. The UI design elements under scrutiny - likes, notifications, in-app messaging - are standard across many platforms including Facebook, Instagram, X, and Zalo. As discussed in Chapter 5, users may migrate to other platforms if TikTok becomes overly restricted. The Government should therefore avoid singling out TikTok. Instead, regulations must apply uniformly across all platforms to ensure fairness and promote consistent safety standards.

In summary, the Government should adopt a combined strategy that integrates strict regulation of social media service providers with comprehensive public education. These measures must be informed by up-to-date scientific research on user behavior and platform design, ensuring that policies remain relevant and effective. Importantly, regulations should not target TikTok in isolation, as the platform's features are widely replicated across the industry. Instead, all social media platforms operating in Vietnam must be held equally accountable to promote uniform safety standards and prevent harmful design practices from being shifted or reproduced elsewhere.

9. Reflection: Applying STAMP and STPA to Unreflective Habit Formation on TikTok

The application of STAMP and STPA to the TikTok system reveals both its potential advantages and notable limitations. These challenges and benefits are discussed below.

Applying STAMP and STPA to unreflective habit formation on TikTok is challenging, primarily because the methodology was originally developed for physical engineering systems. The research encountered two fundamental problems.

Firstly, it is challenging to frame hazards in STAMP terminology for unreflective habit formation, which is a pernicious harm, due to their abstract nature. Unfortunately, the STAMP definitions of accidents and hazards do not encompass the level of abstraction inherent in unreflective habit formation. In STAMP, accidents refer to events resulting in perceivable losses, and hazards refer to system conditions or states that contribute to losses in case of accidents. The accidental event plays a central role in STAMP as it serves as a time pivot to identify the hazardous system states or conditions that contribute to the resulting losses. However, among the three identified losses resulting from unreflective habit formation, it is difficult to pinpoint a specific accidental event that results in the loss of the ability to focus on thought-provoking materials or in identity rigidification. To effectively use the STAMP model to conceptualize these losses, the research suggests defining a pseudo-accidental event, which could take the form of a loss threshold above which the loss becomes unacceptable. Admittedly, identifying this threshold is already a rigorous academic expedition in itself. Until a standardized threshold is established, applying the STAMP model and STPA method to the topic remains highly unfruitful. As long as disputes exist around measurement or acceptance levels of these losses, the analyst must take a normative stance, which can introduce variation into the STPA analysis depending on the stance taken.

Given the complexity in identifying loss thresholds, this research adopts an alternative application of the STAMP model by shifting the focus from losses to the harm of unreflective habit formation itself. Specifically, the research treats each instance of users impulsively using TikTok in response to stable contexts as a hazard to be prevented. Since individual actions are more tangible and observable than abstract, pernicious losses, this approach offers a more manageable way to conceptualize the problem. Based on this shift, there are some slight deviations from the typical STPA methodology. For example, most UCAs, their corresponding controller constraints, and the loss scenarios are linked directly to the hazard rather than to specific losses. However, this deviation does not compromise the validity of the STPA results. Therefore, this approach remains applicable.

Secondly, the categories of causal factors for identifying loss scenarios are not fully applicable to the social media socio-technical system. Three out of the four categories - information

and feedback, control path, and controlled processes - were designed for conventional engineering systems involving actuators, sensors, and physical components. While it is true that TikTok includes physical components such as user devices and TikTok servers, the causal factor categories concerning these components are not entirely relevant in this analysis, because these devices are designed and tested to function in hazardous ways. In other words, as they are inherently hazardous, it is unfruitful to strive to identify hazardous system physical causal factors emerging in their operation. Another obstacle to identifying loss scenarios in STAMP terminology is again the abstract nature of the losses. Without specific accidental events, it is difficult to distinguish causal factors that clearly separate normal from hazardous system behaviors.

On the other hand, the STAMP model and STPA method provide a systematic and methodological approach broad enough to avoid the pitfall of analyzing isolated components while detailed enough to support actionable insights and recommendations.

Firstly, the control structure, comprising the functional control diagram and the hierarchical control structure, provides conceptual models at both the technical and social levels of Tik-Tok's socio-technical system, enabling a comprehensive systemic analysis. The functional control diagram reveals the intricate interplays among different controllers and controlled processes. From this model, detailed UCA are identified and addressed at the component level. As it also exposes the chain of command, it identifies the most influential controller - the development team - at the operational level for ensuring user safety on TikTok. The hierarchical control structure provides complementary insights to the functional control diagram. It reveals higher chains of command within TikTok and society, showing that the development team is not necessarily the root cause when viewed at a higher level. It points to the Vietnam Government as the actor bearing ultimate responsibility for user safety on TikTok.

Secondly, the UCA and controller constraints give concrete details about unsafe system behaviors by mapping specific actions to corresponding controllers and contexts. The four conditions for UCA offer a systematic way to examine all actions, making it harder to overlook hazardous behaviors. Since the focus is placed on actions, controllers, and context, these components can serve as practical checklists to inspect and mitigate unsafe behaviors from the system.

Thirdly, the loss scenarios, which serve as explanations for UCAs, offer strong foundations for specific policy recommendations. Each loss scenario - linking particular system states, contexts, UCAs, and causal factors - can provide a relatively comprehensive narrative of how hazards manifest. In other words, these scenarios can be used as concrete problem framings, which are essential for identifying targeted and actionable resolution strategies.

In summary, while the application of the STAMP model and STPA method to social media introduces several challenges, it also offers substantial benefits. One major challenge lies in the conceptualization stage: due to the abstract and complex nature of the harms involved, traditional loss-centered framing is difficult to implement. This research addresses the issue by shifting the conceptual and analytical focus from losses to the process of unreflective habit formation itself, particularly emphasizing the prevention of repeated impulsive actions in stable contexts. Another challenge is that not all causal analysis categories within STPA are equally applicable to identifying loss scenarios in this domain. Despite these limitations, the approach proves beneficial in several ways. The control structure offers a systematic and holistic overview of the socio-technical system, allowing the problem to be framed in

terms of dynamic interactions rather than isolated components. Additionally, the identification of UCAs and controller constraints produces detailed and functional insights rather than abstract generalizations. Most notably, the development of loss scenarios enables comprehensive problem framings, each of which captures a critical aspect of the overall issue and provides a foundation for specific and actionable policy strategies.

10. Conclusion

Social media leverages neurological reward cycles to reinforce habitual behavior in users (Esposito and Ferreira 2024). User autonomy is infringed in three out of the five dimensions identified in Section 4.3. First, the decision dimension is compromised, as the act of using social media often stems from automatic responses to stimuli, rather than deliberate critical reflection. Second, the execution dimension is undermined due to the absence of intentionality. Third, by fostering a hyper-stimulating system filled with unpredictable rewards, social media creates conditions that are hostile to user autonomy (system hostility). When habits are formed under such conditions, the process can be characterized as unreflective habit formation.

Among Vietnamese university students, TikTok usage becomes habitual without their conscious awareness of this formation process. TikTok presents itself as an altruistic platform that offers entertainment, social connection, information, aspirational content, and opportunities for virality and self-expression. Because of these perceived benefits, users struggle to disengage from the platform and seldom reflect critically on its influence. The underlying neurological reward mechanisms, which TikTok exploits, guide their behaviors in ways that bypass deliberate choice - thereby solidifying their usage into an unreflective habit.

Three major losses emerge from unreflective habit formation. First, users experience a significant loss of time, as TikTok's design makes it difficult to regulate screen time effectively. Although the platform provides screen time management tools, they are largely symbolic rather than functional (Haidt 2023). Second, users lose the ability to concentrate for extended periods. The habit of constantly paying partial attention to TikTok's rapid content stream diminishes their capacity for deep thinking (Norlock 2021). Additionally, TikTok's brevity-oriented design rewires users' cognitive patterns to favor shallow and transient content over sustained, reflective engagement (Carr 2011). Third, highly personalized content feeds limit exposure to diverse personas. When users repeatedly engage with content reflecting idealized versions of themselves, they may begin to internalize and enact those personas offline, gradually becoming their online selves (Attrill-Smith 2019).

The responsibility for regulating social media platforms like TikTok lies primarily with the Vietnam Government. The Vietnam Government should adopt a combined strategy that integrates strict regulation of social media service providers with comprehensive public education. These measures must be informed by up-to-date scientific research on user behavior and platform design, ensuring that policies remain relevant and effective. Importantly, regulations should not target TikTok in isolation, as the platform's features are widely replicated across the industry. Instead, all social media platforms operating in Vietnam must be held equally accountable to promote uniform safety standards and prevent harmful design practices from being shifted or reproduced elsewhere.

In summary, applying the STAMP model and STPA method to social media reveals both challenges and benefits. Conceptualizing abstract harms proved difficult, leading this research to shift focus from losses to the process of unreflective habit formation, emphasizing the

prevention of repeated impulsive use. Some causal categories in STPA were also less relevant for identifying loss scenarios. Despite these challenges, the control structure provides a systematic overview, while UCAs, controller constraints, and loss scenarios offer detailed, actionable insights that support targeted and comprehensive policy strategies.

11. Criticisms and Future Improvements

This research acknowledges several limitations, including its methodological approach, and the application of the STAMP model and STPA method.

Regarding methodology, while interviews offer in-depth insights into individual narratives and allow participants to express themselves, the limited number of interviews challenges the generalizability of the findings. Although consistent patterns emerged across all 12 interviews, it is uncertain whether those patterns can be projected to the broader population. A future improvement would be to expand the participant pool or apply more structured sampling techniques, such as stratified sampling, to control for participant profiles and enable the discovery of more diverse patterns.

Regarding the method used to identify losses from unreflective habit formation, the research relies on existing literature related to the loss of user autonomy. However, it does not provide empirical evidence to substantiate these claims. A potential improvement would involve an in-depth empirical exploration of the losses resulting from diminished user autonomy and unreflective habit formation. This would also help validate the STAMP conceptualization and STPA analysis.

With respect to the selection of the STAMP model and STPA method to explore user safety on TikTok, the granularity of the approach, which focused on specific harms and losses, is both a strength and a limitation. On one hand, this specificity allows for actionable analysis outputs. On the other hand, as the number of harms and losses increases, the workload required to conduct STPA analyses also grows. However, one-size-fits-all solutions rarely address problems at the required level of detail. In the case of TikTok and social media more broadly, such generic solutions may generate ambiguity that platforms like TikTok could exploit to bypass safety requirements.

In the STPA analysis, the hierarchical control structure can be improved in two major ways. First, it is important to identify other influential actors beyond the Vietnam Government and TikTok, such as independent observers or advocacy groups focused on user safety. Although adding more actors increases the complexity of the system, a few additional actors—particularly those that empower users—do not necessarily make the complexity unmanageable. Second, the roles within the TikTok T&S team and their responsibilities in ensuring safe operational processes must be more clearly delineated. Merely reviewing job listings from TikTok is insufficient for this purpose. Rather, interviews with TikTok T&S staff could provide critical insights. Future research should therefore further investigate influential actors promoting user interests and conduct a more thorough examination of the TikTok T&S team to refine the hierarchical control structure.

Bibliography

Abbas, Roba and Katina Michael (2023). "Socio-Technical Theory: A review". In: *TheoryHub Book*. S. Papagiannidis (Ed). ISBN: 978-1-7396044-0-0. URL: https://open.ncl.ac.uk/theory-library/socio-technical-theory.pdf.

Anderson, Ian A. and Wendy Wood (June 2023). "Social motivations' limited influence on habitual behavior: Tests from social media engagement." In: *Motivation Science* 9.2, pp. 107–119. ISSN: 2333-8121, 2333-8113. DOI: 10.1037/mot0000292. URL: https://doi.apa.org/doi/10.1037/mot0000292 (visited on 06/07/2025).

Anker, Thomas Boysen (2024). "Consumer Autonomy and Social Technology: The Case of Social Media Algorithms and the Metaverse". In: *Welcome to The New Normal: Life After The Chaos*. Ed. by Vincent Jeseo and Juliann Allen. Cham: Springer Nature Switzerland, pp. 99–110. ISBN: 978-3-031-49039-2. DOI: 10.1007/978-3-031-49039-2_10.

Attrill-Smith, Alison (May 16, 2019). "The Online Self". In: *The Oxford Handbook of Cyberpsychology*. Ed. by Alison Attrill-Smith et al. Oxford University Press, pp. 16–34. ISBN: 978-0-19-881274-6. DOI: 10.1093/oxfordhb/9780198812746.013.37. URL: https://academic.oup.com/edited-volume/41352/chapter/352512433 (visited on 03/27/2025).

Bao Chinh Phu (Dec. 7, 2023). *TikTok thuc hien cam ket theo ket luat thanh tra cua Bo TT&TT nhu the nao?* baochinhphu.vn. URL: https://baochinhphu.vn/tiktok-thuc-hien-cam-ket-theo-ket-luan-thanh-tra-cua-bo-tttt-nhu-the-nao-102231207183904283.htm (visited on 04/25/2025).

Bennett, Dan et al. (Apr. 19, 2023). "How does HCI Understand Human Agency and Autonomy?" In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. New York, NY, USA: Association for Computing Machinery, pp. 1–18. ISBN: 978-1-4503-9421-5. DOI: 10.1145/3544548.3580651. URL: https://dl.acm.org/doi/10.1145/3544548.3580651 (visited on 04/16/2025).

Boyle, Karen and Susan Berridge (Aug. 25, 2023). *The Routledge Companion to Gender, Media and Violence*. Taylor & Francis. 560 pp. ISBN: 978-1-000-91935-6.

Carr, Nicholas G. (2011). *The shallows: what the Internet is doing to our brains*. Norton pbk. [ed.] New York: W.W. Norton. 280 pp. ISBN: 978-0-393-33975-8.

Choi, Bernard C.K. and Anita W.P. Pak (Dec. 15, 2004). "A Catalog of Biases in Questionnaires". In: *Preventing Chronic Disease* 2.1, A13. ISSN: 1545-1151. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1323316/ (visited on 12/13/2024).

Creswell, John W. (2008). *Research design: qualitative, quantitative, and mixed methods approaches*. 3. ed., [Nachdr.] Los Angeles: SAGE Publ. 260 pp. ISBN: 978-1-4129-6556-9 978-1-4129-6557-6.

DataReportal (2025). *Global Social Media Statistics*. DataReportal – Global Digital Insights. URL: https://datareportal.com/social-media-users (visited on 05/13/2025).

De Souza, Nivio Paula et al. (Nov. 2019). "STAMP-Based Approach to Analyze Safety, Security and Data Privacy". In: *2019 9th Latin-American Symposium on Dependable Computing (LADC)*. 2019 9th Latin-American Symposium on Dependable Computing (LADC). Natal, Brazil: IEEE, pp. 1–10. ISBN: 978-1-7281-6622-3. DOI: 10.1109/LADC48089.2019.8995717. URL: https://ieeexplore.ieee.org/document/8995717/ (visited on 12/16/2024).

Bibliography

Esposito, Fabrizio and Thaís Maciel Cathoud Ferreira (Dec. 2024). "Addictive Design as an Unfair Commercial Practice: The Case of Hyper-Engaging Dark Patterns". In: *European Journal of Risk Regulation* 15.4, pp. 999–1016. ISSN: 1867-299X, 2190-8249. DOI: 10.1017/err.2024.8. URL: https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/addictive-design-as-an-unfair-commercial-practice-the-case-of-hyperengaging-dark-patterns/038CED800E0CAD86EC5B5216E0AA88DD#fn4 (visited on 05/14/2025).

Friedberg, Ivo et al. (June 2017). "STPA-SafeSec: Safety and security analysis for cyber-physical systems". In: *Journal of Information Security and Applications* 34, pp. 183–196. ISSN: 22142126. DOI: 10.1016/j.jisa.2016.05.008. URL: https://linkinghub.elsevier.com/retrieve/pii/S2214212616300850 (visited on 12/16/2024).

Gal, Michal S. (2018). "Algorithmic Challenges to Autonomous Choice". In: *Michigan Technology Law Review* 25, p. 59. URL: https://heinonline.org/HOL/Page?handle=hein.journals/mttlr25&id=65&div=&collection=.

Ghosh, Avijit and Dhanya Lakshmi (Aug. 2, 2023). *Dual Governance: The intersection of centralized regulation and crowdsourced safety mechanisms for Generative AI*. DOI: 10.48550/arXiv.2308.04448. arXiv: 2308.04448[cs]. URL: http://arxiv.org/abs/2308.04448 (visited on 01/08/2025).

Haidt, Jon (Mar. 9, 2023). *TikTok Is Harming Children at an Industrial Scale*. URL: https://www.afterbabel.com/p/industrial-scale-harm-tiktok (visited on 05/21/2025).

Harzing, Anne-Wil (2007). *Publish or Perish*. URL: https://harzing.com/resources/publish-or-perish.

Hata, Akihiro et al. (Jan. 2015). "Using Hazard Analysis STAMP/STPA in Developing Model-Oriented Formal Specification toward Reliable Cloud Service". In: *2015 International Conference on Platform Technology and Service*. 2015 International Conference on Platform Technology and Service (PlatCon). Jeju, South Korea: IEEE, pp. 23–24. ISBN: 978-1-4799-1888-1. DOI: 10.1109/PlatCon.2015.14. URL: http://ieeexplore.ieee.org/document/7079623/ (visited on 12/16/2024).

Huynh, Thach (2025). *TikTok is now the biggest short-video platform in Vietnam*. URL: https://www.decisionlab.co/blog/tiktok-is-now-the-biggest-short-video-platform-in-vietnam (visited on 01/08/2025).

Icenogle, Grace et al. (Feb. 2019). "Adolescents' Cognitive Capacity Reaches Adult Levels Prior to Their Psychosocial Maturity: Evidence for a "Maturity Gap" in a Multinational, Cross-Sectional Sample". In: *Law and human behavior* 43.1, pp. 69–85. ISSN: 0147-7307. DOI: 10.1037/lhb0000315. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6551607/ (visited on 05/14/2025).

Jonsdottir, I. H. et al. (Mar. 1, 2013). "Cognitive impairment in patients with stress-related exhaustion". In: *Stress* 16.2. Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/10253890.2012.708950, pp. 181–190. ISSN: 1025-3890. DOI: 10.3109/10253890.2012.708950. URL: https://doi.org/10.3109/10253890.2012.708950 (visited on 05/17/2025).

Kaplan, Andreas M. and Michael Haenlein (Jan. 1, 2010). "Users of the world, unite! The challenges and opportunities of Social Media". In: *Business Horizons* 53.1, pp. 59–68. ISSN: 0007-6813. DOI: 10.1016/j.bushor.2009.09.003. URL: https://www.sciencedirect.com/science/article/pii/S0007681309001232 (visited on 05/12/2025).

Karizat, Nadia et al. (Oct. 13, 2021). "Algorithmic Folk Theories and Identity: How TikTok Users Co-Produce Knowledge of Identity and Engage in Algorithmic Resistance". In: *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW2), pp. 1–44. ISSN: 2573-0142. DOI: 10.1145/3476046. URL: https://dl.acm.org/doi/10.1145/3476046 (visited on 04/02/2025).

Bibliography

Keenan, Cormac (Apr. 23, 2021). *How our safety teams work to protect our community*. Newsroom — TikTok. URL: https://newsroom.tiktok.com (visited on 04/24/2025).

Klenk, Michael and Jeff Hancock (2019). "Autonomy and Online Manipulation". In: *Internet Policy Review* 1, pp. 1–11.

Kluyver, Thomas et al. (June 1, 2016). *Jupyter Notebooks—a publishing format for reproducible computational workflows*. Pages: 87-90 Publication Title: IOS Press ADS Bibcode: 2016ppap.book...87K. DOI: 10.3233/978-1-61499-649-1-87. URL: https://ui.adsabs.harvard.edu/abs/2016ppap.book...87K (visited on 12/14/2024).

Kubheka, Brenda Z., Vanessa Carter, and Job Mwaura (July 9, 2020). "Social media health promotion in South Africa: Opportunities and challenges". In: *African Journal of Primary Health Care & Family Medicine* 12.1. ISSN: 2071-2936, 2071-2928. DOI: 10.4102/phcfm.v12i1.2389. URL: http://www.phcfm.org/index.php/PHCFM/article/view/2389 (visited on 05/13/2025).

Laacke, Sebastian et al. (July 3, 2021). "Artificial Intelligence, Social Media and Depression. A New Concept of Health-Related Digital Autonomy". In: *The American Journal of Bioethics* 21.7. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/15265161.2020.1863515, pp. 4–20. ISSN: 1526-5161. DOI: 10.1080/15265161.2020.1863515. URL: https://doi.org/10.1080/15265161.2020.1863515 (visited on 05/12/2025).

Le, Viet Tho and Jonathon Hutchinson (2022). "Regulating social media and influencers within Vietnam". In: *Policy & Internet* 14.3. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/poi3.3 pp. 558–573. ISSN: 1944-2866. DOI: 10.1002/poi3.325. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/poi3.325 (visited on 05/14/2025).

Leveson, Nancy G. (Jan. 13, 2012). *Engineering a Safer World: Systems Thinking Applied to Safety*. The MIT Press. ISBN: 978-0-262-29824-7. DOI: 10.7551/mitpress/8179.001.0001. URL: https://direct.mit.edu/books/book/2908/Engineering-a-Safer-WorldSystems-Thinking-Applied (visited on 11/28/2024).

Leveson, Nancy G. and John P. Thomas (Mar. 2018). *STPA Handbook*.

Levin, Sam (May 1, 2017). "Facebook told advertisers it can identify teens feeling 'insecure' and 'worthless'". In: *The Guardian*. ISSN: 0261-3077. URL: https://www.theguardian.com/technology/2017/may/01/facebook-advertising-data-insecure-teens (visited on 05/12/2025).

Lewis, Paul (Oct. 6, 2017). "'Our minds can be hijacked': the tech insiders who fear a smartphone dystopia". In: *The Guardian*. ISSN: 0261-3077. URL: https://www.theguardian.com/technology/2017/oct/05/smartphone-addiction-silicon-valley-dystopia (visited on 05/12/2025).

Mahoney, James and Gary Goertz (2006). "A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research". In: *Political Analysis* 14.3, pp. 227–249. ISSN: 1047-1987, 1476-4989. DOI: 10.1093/pan/mpj017. URL: https://www.cambridge.org/core/product/identifier/S1047198700001418/type/journal_article (visited on 05/15/2025).

Malik, Hanna Maria et al. (Mar. 1, 2022). "Dynamics of Social Harms in an Algorithmic Context". In: *International Journal for Crime, Justice and Social Democracy* 11.1. Number: 1 Publisher: Journal of Crime, Justice and Democracy, pp. 182–195. ISSN: 2202-8005. DOI: 10.5204/ijcjsd.2141. URL: https://www.crimejusticejournal.com/article/view (visited on 04/01/2025).

Marcus, Rachel and Diana Koester (Feb. 5, 2024). *How does social media influence gender norms among adolescent boys?* Align Platform. URL: https://www.alignplatform.org/resources/brief-social-media-gender-norms-adolescent-boys (visited on 01/07/2025).

Marin, Lavinia (Feb. 17, 2025). "Becoming oneself online: Social media platforms as environments for self-transformation". In: *Humanities and Social Sciences Communications* 12.1,

p. 219. ISSN: 2662-9992. DOI: 10.1057/s41599-025-04495-7. URL: https://www.nature.com/articles/s41599-025-04495-7 (visited on 03/27/2025).

Mik, Eliza (Jan. 2, 2016). "The erosion of autonomy in online consumer transactions". In: *Law, Innovation and Technology* 8.1, pp. 1–38. ISSN: 1757-9961, 1757-997X. DOI: 10.1080/17579961.2016.1161893. URL: https://www.tandfonline.com/doi/full/10.1080/17579961.2016.1161893 (visited on 05/12/2025).

Ministry of Information and Communications (Oct. 5, 2023). *Ministry of Information and Communications announces the examination results of TikTok in Vietnam*. Ministry of Information and Communications — Authority of Broadcasting and Electronic Information. URL: https://abei.gov.vn/thong-tin-dien-tu/bo-thong-tin-va-truyen-thong-thong-bao-ket-qua-kiem-tra-toan-dien-hoat-dong-cua-tiktok-tai-viet-nam-/108209 (visited on 12/19/2024).

Ministry of Science and Technology (Apr. 29, 2023). *Vi pham cua mang xa hoi Tiktok tai Viet Nam*. Cong Thong tin dien tu Bo Khoa hoc va Cong nghe. URL: https://mst.gov.vn/vi-pham-cua-mang-xa-hoi-tiktok-tai-viet-nam-197158512.htm (visited on 04/27/2025).

Nguyen, Tham Thi et al. (Dec. 2025). "Patterns of social media use in Vietnamese youths: status and associations with psychological well-being". In: *BMC Public Health* 25.1. Number: 1 Publisher: BioMed Central, pp. 1–11. ISSN: 1471-2458. DOI: 10.1186/s12889-025-22337-8. URL: https://link-springer-com.tudelft.idm.oclc.org/article/10.1186/s12889-025-22337-8 (visited on 05/14/2025).

Norlock, Kathryn (2021). "Free and Always Will Be? On Social Media Participation as it Undermines Individual Autonomy". In: *Canadian Journal of Practical Philosophy* 5.1. Number: 1. URL: https://ojs.uwindsor.ca/index.php/cjpp/article/view/8195 (visited on 05/12/2025).

Obar, Jonathan A. and Steve Wildman (Oct. 1, 2015). "Social media definition and the governance challenge: An introduction to the special issue". In: *Telecommunications Policy*. SPECIAL ISSUE ON THE GOVERNANCE OF SOCIAL MEDIA 39.9, pp. 745–750. ISSN: 0308-5961. DOI: 10.1016/j.telpol.2015.07.014. URL: https://www.sciencedirect.com/science/article/pii/S0308596115001172 (visited on 05/12/2025).

Park, Kristy et al. (Mar. 2023). *The impact of the use of social media on women and girls*. URL: https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2023)743341 (visited on 06/14/2025).

Patriarca, Riccardo et al. (Feb. 1, 2022). "The past and present of System-Theoretic Accident Model And Processes (STAMP) and its associated techniques: A scoping review". In: *Safety Science* 146, p. 105566. ISSN: 0925-7535. DOI: 10.1016/j.ssci.2021.105566. URL: https://www.sciencedirect.com/science/article/pii/S0925753521004082 (visited on 12/01/2024).

Pereira, Daniel et al. (2017). "Towards Combined Safety and Security Constraints Analysis". In: *Computer Safety, Reliability, and Security*. Ed. by Stefano Tonetta, Erwin Schoitsch, and Friedemann Bitsch. Vol. 10489. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 70–80. ISBN: 978-3-319-66283-1 978-3-319-66284-8. DOI: 10.1007/978-3-319-66284-8_7. URL: http://link.springer.com/10.1007/978-3-319-66284-8_7 (visited on 12/16/2024).

Pinder, Charlie et al. (June 12, 2018). "Digital Behaviour Change Interventions to Break and Form Habits". In: *ACM Trans. Comput.-Hum. Interact.* 25.3, 15:1–15:66. ISSN: 1073-0516. DOI: 10.1145/3196830. URL: https://dl.acm.org/doi/10.1145/3196830 (visited on 06/08/2025).

Porcelli, Anthony J and Mauricio R Delgado (Apr. 1, 2017). "Stress and decision making: effects on valuation, learning, and risk-taking". In: *Current Opinion in Behavioral Sci-*

ences. Stress and behavior 14, pp. 33–39. ISSN: 2352-1546. DOI: `10.1016/j.cobeha.2016.11.015`. URL: `https://www.sciencedirect.com/science/article/pii/S2352154616301905` (visited on 05/17/2025).

Sahebi, Siavosh and Paul Formosa (Sept. 1, 2022). "Social Media and its Negative Impacts on Autonomy". In: *Philosophy & Technology* 35.3. Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 3 Publisher: Springer Netherlands, pp. 1–24. ISSN: 2210-5441. DOI: `10.1007/s13347-022-00567-7`. URL: `https://link-springer-com.tudelft.idm.oclc.org/article/10.1007/s13347-022-00567-7` (visited on 04/01/2025).

Savolainen, Laura and Minna Ruckenstein (June 1, 2024). "Dimensions of autonomy in human–algorithm relations". In: *New Media & Society* 26.6. Publisher: SAGE Publications, pp. 3472–3490. ISSN: 1461-4448. DOI: `10.1177/14614448221100802`. URL: `https://doi.org/10.1177/14614448221100802` (visited on 04/01/2025).

Schlosser, Markus (2019). "Agency". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2019. Metaphysics Research Lab, Stanford University. URL: `https://plato.stanford.edu/archives/win2019/entries/agency/` (visited on 04/07/2025).

Shapiro, Stuart S. (May 2016). "Privacy Risk Analysis Based on System Control Structures: Adapting System-Theoretic Process Analysis for Privacy Engineering". In: *2016 IEEE Security and Privacy Workshops (SPW)*. 2016 IEEE Security and Privacy Workshops (SPW). San Jose, CA: IEEE, pp. 17–24. ISBN: 978-1-5090-3690-5. DOI: `10.1109/SPW.2016.15`. URL: `http://ieeexplore.ieee.org/document/7527748/` (visited on 12/16/2024).

Shelby, Renee et al. (Aug. 8, 2023). "Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction". In: *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. AIES '23: AAAI/ACM Conference on AI, Ethics, and Society. Montr\'{e}al QC Canada: ACM, pp. 723–741. ISBN: 979-8-4007-0231-0. DOI: `10.1145/3600211.3604673`. URL: `https://dl.acm.org/doi/10.1145/3600211.3604673` (visited on 02/05/2025).

Sidhu, Amardeep and Allen Moulton (Apr. 14, 2020). "Enriching Systems Theory Based Cyber-Security and Safety Analysis Using Stakeholder Value Networks". In: *SAE International Journal of Advances and Current Practices in Mobility* 2.6, pp. 3083–3092. ISSN: 2641-9645. DOI: `10.4271/2020-01-0143`. URL: `https://www.sae.org/content/2020-01-0143` (visited on 12/16/2024).

Snyder, Hannah (Nov. 1, 2019). "Literature review as a research methodology: An overview and guidelines". In: *Journal of Business Research* 104, pp. 333–339. ISSN: 0148-2963. DOI: `10.1016/j.jbusres.2019.07.039`. URL: `https://www.sciencedirect.com/science/article/pii/S0148296319304564` (visited on 12/13/2024).

Statista (Feb. 2025a). *Biggest social media platforms by users 2025*. Statista. URL: `https://www-statista-com.tudelft.idm.oclc.org/statistics/272014/global-social-networks-ranked-by-number-of-users/` (visited on 05/13/2025).

— (Feb. 2025b). *Internet and social media users in the world 2025*. Statista. URL: `https://www-statista-com.tudelft.idm.oclc.org/statistics/617136/digital-population-worldwide/` (visited on 05/13/2025).

Stoljar, Natalie (May 2, 2013). "Feminist Perspectives on Autonomy". In: Last Modified: 2024-05-20. URL: `https://plato-stanford-edu.tudelft.idm.oclc.org/entries/feminism-autonomy/?utm_source=chatgpt.com` (visited on 04/04/2025).

Stone, Linda (Nov. 30, 2009). *Beyond Simple Multi-Tasking: Continuous Partial Attention*. Linda Stone. URL: `https://lindastone.net/2009/11/30/beyond-simple-multi-tasking-continuous-partial-attention/` (visited on 05/17/2025).

Susser, Daniel, Beate Roessler, and Helen Nissenbaum (June 30, 2019). *Technology, autonomy, and manipulation*. Publisher: Alexander von Humboldt Institute for Internet and Society

gGmbH. DOI: [10.14763/2019.2.1410](https://policyreview.info/articles/analysis/technology-autonomy-and-manipulation). URL: https://policyreview.info/articles/analysis/technology-autonomy-and-manipulation (visited on 05/02/2025).

Tarafdar, Pratik and Indranil Bose (July 3, 2019). "Systems theoretic process analysis of information security: the case of aadhaar". In: *Journal of Organizational Computing and Electronic Commerce* 29.3, pp. 209–222. ISSN: 1091-9392, 1532-7744. DOI: [10.1080/10919392.2019.1598608](https://www.tandfonline.com/doi/full/10.1080/10919392.2019.1598608). URL: https://www.tandfonline.com/doi/full/10.1080/10919392.2019.1598608 (visited on 12/16/2024).

Thu Hang (Nov. 6, 2023). *TikTok cam ket xoa bo tai khoan tre em duoi 13 tuoi, bao ve tre em*. thanhnien.vn. Section: Thoi su. URL: https://thanhnien.vn/tiktok-cam-ket-xoa-bo-tai-khoan-tre-em-duoi-13-tuoi-bao-ve-tre-em-185231106194906007.htm (visited on 04/25/2025).

Thu Vien Phap Luat (Jan. 10, 2025a). *Luat an ninh mang 2018 so 24/2018/QH14 ap dung 2025 moi nhat*. Thu Vien Phap Luat. URL: https://thuvienphapluat.vn/van-ban/Cong-nghe-thong-tin/Luat-an-ninh-mang-2018-351416.aspx (visited on 04/24/2025).

— (Mar. 22, 2025b). *Nghi dinh 147/2024/NĐ-CP quan ly cung cap su dung dich vu Internet, thong tin tren mang moi nhat*. Thu Vien Phap Luat. URL: https://thuvienphapluat.vn/van-ban/Cong-nghe-thong-tin/Nghi-dinh-147-2024-ND-CP-quan-ly-cung-cap-su-dung-dich-vu-Internet-thong-tin-tren-mang-480755.aspx (visited on 04/24/2025).

— (Apr. 23, 2025c). *Quyet dinh 874/QĐ-BTTTT 2021 Bo Quy tac ung xu tren mang xa hoi*. Thu Vien Phap Luat. URL: https://thuvienphapluat.vn/van-ban/Cong-nghe-thong-tin/Quyet-dinh-874-QD-BTTTT-2021-Bo-Quy-tac-ung-xu-tren-mang-xa-hoi-478154.aspx (visited on 04/24/2025).

TikTok (July 26, 2021). *One Year in Trust & Safety: Lessons from Kyung-Hoon Kim*. URL: https://lifeattiktok.com/blog/detail/6989168054970763534? (visited on 04/24/2025).

— (Jan. 12, 2023). *Grow With TikTok: A Career in Trust and Safety*. URL: https://careers.tiktok.com/blog/detail/7187516805735631161 (visited on 03/02/2025).

— (Apr. 17, 2024). *Community Guidelines*. URL: https://www.tiktok.com/community-guidelines/en/overview (visited on 04/25/2025).

— (2025a). *About — TikTok*. TikTok. URL: https://www.tiktok.com/about (visited on 05/14/2025).

— (2025b). *AIGC Search Safety Product Manager, Trust & Safety*. URL: https://lifeattiktok.com/search/7489316022575728903 (visited on 04/24/2025).

— (2025c). *Content Moderator (Feature) - VN*. URL: https://lifeattiktok.com/search/7347543610411469065 (visited on 02/25/2025).

— (2025d). *Data Scientist, Trust & Safety*. URL: https://lifeattiktok.com/search/7491479659700553991 (visited on 04/24/2025).

— (2025e). *Frontend Software Engineer (Data Insights) , Trust and Safety Platform*. URL: https://lifeattiktok.com/search/7454060448854329618 (visited on 05/29/2025).

— (2025f). *Head of Risk Prevention APAC, Trust & Safety*. URL: https://lifeattiktok.com/search/7485682709625932040 (visited on 05/29/2025).

— (2025g). *Head of Trust & Safety Operations Quality AMS*. URL: https://lifeattiktok.com/search/7335135620324739366 (visited on 03/02/2025).

— (2025h). *High Harm Policy Manager, Bangladesh, Trust & Safety*. URL: https://lifeattiktok.com/search/7496011281888053511 (visited on 05/29/2025).

— (2025i). *How TikTok recommends content — TikTok Help Center*. URL: https://support.tiktok.com/en/using-tiktok/exploring-videos/how-tiktok-recommends-content (visited on 02/25/2025).

— (2025j). *Incident Manager, Trust & Safety (Thai Speaking)*. URL: https://lifeattiktok.com/search/7491933821141010696 (visited on 05/29/2025).

TikTok (2025k). *Integrity Manager, Trust & Safety*. URL: https://lifeattiktok.com/search/7501218385688561928 (visited on 05/29/2025).

— (2025l). *Job Search*. URL: https://lifeattiktok.com/search?recruitment_id_list=1&keyword=&limit=12&job_category_id_list=6709824272505768200%2C6704215913488451847%2C6709824272514156812%2C6704215882479962371%2C6704215901438216462%2C6704215864629004552%2C6704215862603155720&location_code_list=CT_163%2CCT_244%2CCT_169%2CCT_65%2CCT_60%2CCT_97%2CCT_34%2CCT_98%2CCT_134%2CCT_101459%2CCT_1102494%2CCT_1102314%2CCT_101457&subject_id_list=&offset=0 (visited on 04/27/2025).

— (2025m). *Language Quality Assurance (Vietnamese Speaker) - Trust & Safety*. URL: https://lifeattiktok.com/search/7373553963235723570 (visited on 05/29/2025).

— (2025n). *Legal Policy Operations: Child Safety Specialist - Trust & Safety*. URL: https://lifeattiktok.com/search/7478633526716549384 (visited on 05/29/2025).

— (2025o). *Machine Learning Engineer - Trust and Safety, NLP (Australia)*. URL: https://lifeattiktok.com/search/7262985689515723065 (visited on 05/29/2025).

— (2025p). *Product Manager (Risk Investigation & Intervention) - Trust and Safety*. URL: https://lifeattiktok.com/search/7189430898005641509 (visited on 05/29/2025).

— (2025q). *Product Manager, Safety Cloud, Trust & Safety*. URL: https://lifeattiktok.com/search/7482664928482953479 (visited on 04/24/2025).

— (2025r). *Product Manager, Video Moderation, Trust & Safety*. URL: https://lifeattiktok.com/search/7509059644356020498 (visited on 05/29/2025).

— (2025s). *Project Manager - AI Data Operation (Safety), Trust and Safety - Singapore*. URL: https://lifeattiktok.com/search/7275210222100171064 (visited on 05/29/2025).

— (2025t). *Regional Analysis and Prevention Lead (APAC), Trust & Safety*. URL: https://lifeattiktok.com/search/7480533164307089671 (visited on 04/24/2025).

— (2025u). *Tech Lead Engineering- Trust and Safety - Singapore*. URL: https://lifeattiktok.com/search/7450661269201455378 (visited on 05/29/2025).

Timms, Ryan and David Spurrett (Jan. 2, 2023). "Hostile Scaffolding". In: *Philosophical Papers* 52.1, pp. 53–82. ISSN: 0556-8641, 1996-8523. DOI: 10.1080/05568641.2023.2231652. URL: https://www.tandfonline.com/doi/full/10.1080/05568641.2023.2231652 (visited on 06/07/2025).

Wang, Yang, Jasmin Ramadani, and Stefan Wagner (2017). "An Exploratory Study on Applying a Scrum Development Process for Safety-Critical Systems". In: *Product-Focused Software Process Improvement*. Ed. by Michael Felderer et al. Vol. 10611. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 324–340. ISBN: 978-3-319-69925-7 978-3-319-69926-4. DOI: 10.1007/978-3-319-69926-4_23. URL: http://link.springer.com/10.1007/978-3-319-69926-4_23 (visited on 12/16/2024).

Zahle, Julie (Dec. 10, 2024). "Bias and Debiasing Strategies in Qualitative Data Collection". In: *Philosophy of Science*, pp. 1–18. ISSN: 0031-8248, 1539-767X. DOI: 10.1017/psa.2024.70. URL: https://www.cambridge.org/core/journals/philosophy-of-science/article/bias-and-debiasing-strategies-in-qualitative-data-collection/ADD91B2D6B2E9E436EF66379A4D25362?utm_source=chatgpt.com (visited on 05/02/2025).

Zhang, Yingyu et al. (Aug. 1, 2022). "Systems theoretic accident model and process (STAMP): A literature review". In: *Safety Science* 152, p. 105596. ISSN: 0925-7535. DOI: 10.1016/j.ssci.2021.105596. URL: https://www.sciencedirect.com/science/article/pii/S0925753521004367 (visited on 12/04/2024).

A. Interview Codebook

Table A.1.: Interview Codebook

Code	Description	Category
Funny/Entertaining Videos	Including conversational jokes and teasing, humorous quotes, daily vlogs.	Content Genre
Challenges/Trends	Often involving dancing to short cuts of popular songs, hand-only choreography, or viral trends. It becomes a trend when many people replicate the choreography or actions and re-use the soundtrack from the original video.	Content Genre
Food-Related Content	Including cooking videos, food reviews, restaurant/coffee shop reviews, mukbang activities.	Content Genre
Fashion and Clothing	Featuring styling, outfit ideas, vintage styles, reviews, styling tips and guidance.	Content Genre
Beauty and Makeup	Including makeup tutorials, skincare routines, and beauty product reviews.	Content Genre
Music-Related Content	Including music videos, trending songs, and covers of trending songs.	Content Genre
Drama/Viral Events/News	Including recent dramas, scandals, and news updates.	Content Genre
Educational/Informational Content	Including history, marketing knowledge, study tips, inspiring quotes, and book recommendations.	Content Genre
Celebrity/Idol Content	Featuring K-Pop idols, beauty queens, KOLs (Key Opinion Leaders), dancers, product reviewers.	Content Genre
Pet-Related Content	Videos about pets.	Content Genre
Religious Content	Videos related to religious topics.	Content Genre
Game Show Clips	Short clips from popular game shows.	Content Genre
Entertainment and Fun	To laugh, enjoy funny videos, and pass the time.	Motivations for Using TikTok
Relaxation and De-stress	To unwind after school or work.	Motivations for Using TikTok
Learning and Information	To learn new skills, get news updates, find product reviews, and gain knowledge related to their interests or studies.	Motivations for Using TikTok

Table A.1.: Interview Codebook

Code	Description	Category
Following Trends and Feeling Included (FOMO)	To participate in popular dances or trends.	Motivations for Using TikTok
Inspiration and Aspiration	To admire creators' beauty, style, and learn how to improve their own appearance through fashion and beauty routine tips.	Motivations for Using TikTok
Social Connection and Interaction	To connect with friends, maintain streaks, share videos, see what others are doing, interact with their friends' TikTok contents.	Motivations for Using TikTok
Boredom Relief	To occupy themselves when feeling bored.	Motivations for Using TikTok
Habitual Use	Simply out of routine.	Motivations for Using TikTok
Discovering Recommendations	To see what TikTok recommends and discover new contents, such as new coffee shops or restaurants to hang out with friends.	Motivations for Using TikTok
Liking (Hearting) Videos	To show appreciation and support creators, to retrieve videos later when needed.	Engagement with Content
Saving Videos	To watch later, to retrieve when needed (e.g., trends, product information), or increase interaction scores for idols.	Engagement with Content
Leaving Comments	To express opinions, compliment creators, ask questions, or interact with other viewers.	Engagement with Content
Sharing Videos	To send to friends on TikTok or other platforms, or to maintain streaks.	Engagement with Content
Following Creators	To see more content from individuals or accounts they like.	Engagement with Content
Skipping Videos	To avoid contents they are not interested in or find repetitive.	Engagement with Content
Replying to Comments	Some participants, like teachers interacting with students, reply to comments.	Engagement with Content
Reposting Videos	To share content on their own profile, to retrieve later when needed.	Engagement with Content
Admiration for Talent and Skills	Impressed by dancing, singing, musical abilities, and video creation skills.	Interactions with Creators
Admiration for appearance and self-expression	Impressed by appearance, self-expression	Interactions with Creators

Table A.1.: Interview Codebook

Code	Description	Category
Appreciation for Authenticity and Reliability	Preferring creators who are honest in their reviews and share genuine experiences.	Interactions with Creators
Interest in Lifestyle and Style	Following creators to learn about their dressing styles, makeup and skincare routines, and overall vibe.	Interactions with Creators
Influence on Purchasing Decisions	Trusting reviews from certain creators and being persuaded to buy products they promote.	Interactions with Creators
Awareness of Promotional Intent	Recognizing when creators are sponsored or receiving commissions.	Interactions with Creators
Dislike for Certain Styles	Finding overly revealing or 'sexy' content from some creators repulsive.	Interactions with Creators
Following Each Other	Follow accounts to see each other's content.	Interactions with Other Users (Friends)
Tagging in Comments	Sharing interesting or relevant videos and initiating discussions.	Interactions with Other Users (Friends)
Sharing Videos Directly	Using the share function to send videos to friends, especially to keep streak.	Interactions with Other Users (Friends)
Creating Videos Together	Collaborating on trends or funny content.	Interactions with Other Users (Friends)
Discussing TikTok Content Offline	Talking about viral dramas, beauty queens, trends, and recommended places for hanging out.	Interactions with Other Users (Friends)
Maintaining ""Streaks""	A feature that encourages daily messaging to build a consecutive day count, often seen as a testament of friendship.	Interactions with Other Users (Friends)
Nudging/Influencing Each Other	Friends sending videos to encourage adopting certain styles.	Interactions with Other Users (Friends)
Discovering Products Through Videos	Seeing clothing, beauty products, and other items featured in creators' content.	Shopping Behavior
Clicking on Product Links	Using links in video descriptions to access product pages on TikTok Shop or other e-commerce platforms.	Shopping Behavior
Searching for Products on TikTok Shop	Using keywords or image search to find specific items.	Shopping Behavior
Comparing Prices	Checking prices on TikTok Shop and other platforms like Shopee.	Shopping Behavior
Considering product fit	consider if product suits body shape, skin type.	Shopping Behavior
Reviewing financial situation	reflect current financial status, reflect purchases in the month, reviewing if the price is affordable	Shopping Behavior
Reading Reviews and Feedback	Examining comments from previous buyers on product pages.	Shopping Behavior

Table A.1.: Interview Codebook

Code	Description	Category
Saving Items to Basket or ""Save for Later""	To consider purchasing at a later time, or to retrieve the product for future purchase.	Shopping Behavior
Being Influenced by Repetitive Content	Seeing the same products repeatedly after interacting with them, leading to a feeling of being 'nudged' to buy.	Shopping Behavior
Impulse Purchases	Buying items they immediately like, especially if they have been searching for them or if they are on sale.	Shopping Behavior
Seeking Opinions from Friends	Asking friends for advice before buying uncertain items.	Shopping Behavior
Regret After Purchase (or Not Purchasing)	Experiencing conflicting feelings about spending money.	Shopping Behavior
Livestream Shopping	Watching live sessions where creators promote and sell products.	Shopping Behavior
Positive emotions with high video engagement	feeling happy about high interaction with their videos	Feelings and Emotions
Feeling appreciative of learning contents	feeling appreciative of learning contents on TikTok	Feelings and Emotions
Feeling relaxed	when viewing funny and entertaining videos	Feelings and Emotions
Feeling connected with friends	feeling connected with friends as they maintain streak	Feelings and Emotions
Feeling pressured about streaks	feeling pressured by friends to maintain streak	Feelings and Emotions
Feeling motivated about streaks	feeling motivated to maintain streaks with friends	Feelings and Emotions
Feeling fed-up with repetition	when seeing too many repetitive contents	Feelings and Emotions
Feeling embarrassed about flopped contents	when contents are flopped	Feelings and Emotions
Negative emotions when app is inaccessible	sadness, lost, don't know what to do	Feelings and Emotions
Feeling sad by sad contents	participants can feel sad as they view sad contents	Feelings and Emotions
Feeling repulsive	feeling repulsive towards women dressing sexy or overly revealing	Feelings and Emotions
Neutral Emotions	Feeling normal, neither positive nor negative, about TikTok in general.	Feelings and Emotions
Desire and Want	Wanting to buy products, look like creators, or copy trends.	Feelings and Emotions

Table A.1.: Interview Codebook

Code	Description	Category
Intrigue	Feeling curious and wanting to know more about certain products.	Feelings and Emotions
Cuteness Overload	Feeling overwhelmed by cute content.	Feelings and Emotions
Feeling FOMO	feeling FOMO when viewing too many trendy videos	Feelings and Emotions
During Free Time	Using TikTok whenever they have spare moments.	Usage Patterns and Context
Between Classes/Study Breaks	Using it as a short break during school or study sessions.	Usage Patterns and Context
Before Bed	A common time for using TikTok.	Usage Patterns and Context
At Home	Frequently using TikTok while at home.	Usage Patterns and Context
When Out with Friends	Sometimes using TikTok when socializing.	Usage Patterns and Context
At Coffee Shops	A common setting for using TikTok, often for making videos or discussing content.	Usage Patterns and Context
During Meals	Using TikTok while eating.	Usage Patterns and Context
When Feeling Bored	Turning to TikTok to alleviate boredom.	Usage Patterns and Context
Multitasking	Some participants use autoplay to browse while doing other activities.	Usage Patterns and Context
Privacy Settings	Setting accounts to private, controlling who can view posts or profile visits.	Awareness and Use of Settings
Filters	Frequently using filters for videos and photos to enhance appearance or create funny effects.	Awareness and Use of Settings
""Save for Later""/Saved Videos	Using this feature to store videos of interest.	Awareness and Use of Settings
Watch History	Utilizing the history to revisit previously viewed videos.	Awareness and Use of Settings
Repost Feature	Sharing videos on their own profile.	Awareness and Use of Settings
Blocking Unwanted Content (Through Skipping)	Manually avoiding videos they dislike.	Awareness and Use of Settings
Limited Awareness of Other Settings	Many participants were unaware of features like keyword filters, sleep reminders, screen time limits, and screen time breaks.	Awareness and Use of Settings
Deleting Data/Freeing Up Storage	Some users manage storage by deleting TikTok data.	Awareness and Use of Settings

Table A.1.: Interview Codebook

Code	Description	Category
Auto-Play	Some users enable automatic scrolling of videos.	Awareness and Use of Settings
Generally Positive	Viewing it as a source of entertainment, learning, and connection.	Opinions and Perceptions of TikTok
Generally Neutral	Seeing both positive and negative aspects or not having strong feelings either way.	Opinions and Perceptions of TikTok
Awareness of Addictive Potential	Recognizing the tendency to spend too much time on the app.	Opinions and Perceptions of TikTok
Perception of Algorithm	Believing TikTok can ""read their mind"" and shows content based on their interests and interactions.	Opinions and Perceptions of TikTok
Concerns about Negative Content	Disliking inappropriate, repulsive, or misleading content.	Opinions and Perceptions of TikTok
Value for Information and Trends	Appreciating TikTok for providing quick updates on news and trends.	Opinions and Perceptions of TikTok
Importance of Personalization	Liking that TikTok recommends content aligned with their interests.	Opinions and Perceptions of TikTok
Acceptance of Repetition	Although participants are annoyed by repetition, they accept it.	Opinions and Perceptions of TikTok

Colophon

This document was typeset using LaTeX, using the KOMA-Script class scrbook. The main font is Palatino.