



IntechOpen

The Evolving Landscape of Online Identity

Recent Studies and Insights

Edited by Livinus Obiora Nweke



The Evolving Landscape of Online Identity - Recent Studies and Insights

Edited by Livinus Obiora Nweke

Published in London, United Kingdom

The Evolving Landscape of Online Identity – Recent Studies and Insights

<http://dx.doi.org/10.5772/intechopen.1008006>

Edited by Livinus Obiora Nweke

Contributors

Angelo Compierchio, Claudio Santoro, Jennifer Laffier, Livinus Obiora Nweke, Madison Westley, Phillip Tretten, Tooraj Nikoubin, Yogeswar Reddy Thota

© The Editor(s) and the Author(s) 2026

The rights of the editor(s) and the author(s) have been asserted in accordance with the Copyright, Designs and Patents Act 1988. All rights to the book as a whole are reserved by INTECHOPEN LIMITED. The book as a whole (compilation) cannot be reproduced, distributed or used for commercial or non-commercial purposes without INTECHOPEN LIMITED's written permission. Enquiries concerning the use of the book should be directed to INTECHOPEN LIMITED rights and permissions department (permissions@intechopen.com).

Violations are liable to prosecution under the governing Copyright Law.



Individual chapters of this publication are distributed under the terms of the Creative Commons Attribution 4.0 License which permits commercial use, distribution and reproduction of the individual chapters, provided the original author(s) and source publication are appropriately acknowledged. If so indicated, certain images may not be included under the Creative Commons license. In such cases users will need to obtain permission from the license holder to reproduce the material. More details and guidelines concerning content reuse and adaptation can be found at <http://www.intechopen.com/copyright-policy.html>.

Notice

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

First published in London, United Kingdom, 2026 by IntechOpen

IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, 167-169 Great Portland Street, London, W1W 5PF, United Kingdom

For EU product safety concerns: IN TECH d.o.o., Prolaz Marije Krucifikse Kozulić 3, 51000 Rijeka, Croatia, info@intechopen.com or visit our website at intechopen.com.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

The Evolving Landscape of Online Identity – Recent Studies and Insights

Edited by Livinus Obiora Nweke

p. cm.

Print ISBN 978-0-85466-581-5

Online ISBN 978-0-85466-580-8

eBook (PDF) ISBN 978-0-85466-582-2

If disposing of this product, please recycle the paper responsibly.

Meet the editor



Dr. Livinus Obiora Nweke is an Associate Professor at Noroff University College, Norway, and the University of South-Eastern Norway (USN), Norway. He holds a Ph.D. in Information Security and Communication Technology from the Norwegian University of Science and Technology (NTNU), Norway, and an M. Sc. in Computer Science (*summa cum laude*) from Sapienza University of Rome, Italy. A Certified Information Systems

Security Professional (CISSP), he bridges the gap between academic research and practical cybersecurity. His work focuses on infrastructure security, digital assurance, security and risk management, privacy and data protection, artificial intelligence, and information security. Dr. Nweke has published widely in peer-reviewed journals and continues to explore the intersection of technology, security, and ethics in advancing trustworthy digital ecosystems.

Contents

Preface	IX
Chapter 1	1
Introductory Chapter: Rethinking Online Identity at the Human-AI-Edge Nexus <i>by Livinus Obiora Nweke</i>	
Chapter 2	7
Dyadic Influences on the Psyche of AI Online to AI Onlife <i>by Angelo Compierchio, Claudio Santoro and Phillip Tretten</i>	
Chapter 3	19
‘Digital Identity Congruence’: Exploring the Importance of Congruent Online Identities for Adolescents <i>by Jennifer Laffier and Madison Westley</i>	
Chapter 4	43
EdgeAI with TinyML: Redefining Privacy and Security in Online Identity Management <i>by Yogeswar Reddy Thota and Tooraj Nikoubin</i>	
Chapter 5	67
Artificial Intelligence, Trust, and the Architecture of Online Identity: Security Implications in an Algorithmic Era <i>by Livinus Obiora Nweke</i>	

Preface

In an era where artificial intelligence (AI), ubiquitous connectivity, and algorithmic mediation shape how we communicate, transact, and even perceive ourselves, the concept of online identity has become increasingly complex. *The Evolving Landscape of Online Identity – Recent Studies and Insights* brings together perspectives from psychology, education, engineering, and cybersecurity to illuminate how intelligent systems are reshaping digital identity.

This volume emerges from a shared recognition among scholars and practitioners that identity today is not static; it is continuously negotiated between humans and machines, influenced by algorithms, data infrastructures, and social architectures. Each chapter in this collection offers a distinct yet complementary lens on this ongoing transformation, providing readers with a comprehensive view of the psychological, technological, and ethical forces shaping the digital self.

The book opens with an Introductory Chapter.

The second chapter, *Dyadic Influences on the Psyche of AI Online to AI Onlife*, by Angelo Compierchio, Claudio Santoro and Phillip Tretten explores how AI-mediated interfaces, cues, and symbolic environments subtly reconfigure human perception and emotional equilibrium. Their analysis situates identity not only as information, but also as a lived experience, demonstrating how design choices can impact psychological well-being and autonomy.

In the third chapter, ‘*Digital Identity Congruence*’: *Exploring the Importance of Congruent Online Identities for Adolescents*, Jennifer Laffier and Madison Westley extend Carl Rogers’ humanistic notion of congruence to the digital realm. They demonstrate how alignment between young people’s online and offline selves can foster mental health, authenticity, and resilience. Through frameworks such as Positive Youth Development and social-emotional learning, the authors provide actionable insights for educators and platform designers seeking to nurture reflective and authentic digital participation.

The fourth chapter, *EdgeAI with TinyML: Redefining Privacy and Security in Online Identity Management*, by Yogeswar Reddy Thota and Tooraj Nikoubin, shifts the focus to the technological foundations of digital identity. The authors examine how emerging paradigms, such as EdgeAI and Tiny Machine Learning, enable privacy-preserving identity verification directly on devices. They present practical pathways, including WebAuthn/FIDO2 passkeys, trusted execution environments, and federated learning, that bring intelligence closer to the user while reducing exposure of sensitive data. At the same time, they highlight the evolving risks inherent in decentralized and on-device AI, emphasizing the need for continuous security assurance.

The fifth chapter, *Artificial Intelligence, Trust, and the Architecture of Online Identity: Security Implications in an Algorithmic Era*, by Livinus Obiora Nweke, integrates governance, risk, and design perspectives. It examines how AI transforms traditional identity architectures (centralized, federated, and self-sovereign) by expanding attack surfaces and complicating oversight and management. The chapter proposes a comprehensive security-by-design framework that aligns technical controls with regulatory frameworks such as the General Data Protection Regulation (GDPR), the EU Artificial Intelligence Act, and the NIST AI Risk Management Framework. By combining policy vision with practical mechanisms, model cards, human-in-the-loop oversight, and adversarial testing, the author articulates a blueprint for accountable and transparent identity ecosystems.

Collectively, these chapters present a multifaceted portrayal of online identity, encompassing both technological constructs and human experiences. They reveal that sustaining trust in digital identity requires more than cryptographic robustness; it demands ethical design, psychological insight, and governance that bridges innovation and accountability.

This book aims to serve as a reference for researchers, educators, and practitioners seeking to understand and shape the future of identity in an AI-driven world. It invites readers to reflect on a profound challenge of our time: as we engineer systems that recognize us, we must also ensure they respect and empower us.

Livinus Obiora Nweke
Noroff University College and the University of South-Eastern Norway,
Oslo, Norway

Introductory Chapter: Rethinking Online Identity at the Human-AI-Edge Nexus

Livinus Obiora Nweke

1. Setting the stage

In the digital age, identity has evolved far beyond the static notion of a username, password, or government-issued credential [1, 2]. Today, online identity is a dynamic, data-driven construct: continuously inferred, negotiated, and reconstructed through algorithms, interactions, and infrastructures. Each digital trace, biometric gesture, or behavioral cue contributes to a composite “self” that is both human and machine-interpreted. This transformation compels us to ask fundamental questions: Who defines us online? How is trust assigned? and What are the consequences when algorithmic systems mediate our sense of authenticity and belonging?

Online identity now sits at a critical junction of social life and technical systems. It is increasingly embedded in everyday participation, including education, employment, finance, healthcare, and public services. As a result, identity is no longer only about authentication and authorization. It is also about autonomy, assurance, and alignment between human values and intelligent infrastructures. The core challenge is not merely to secure identity, but to preserve legitimacy, inclusiveness, and dignity under expanding capabilities and risks.

A helpful way to approach the topic is to treat identity as an ecosystem of claims, evidence, decisions, and accountability. Claims about “who someone is” or “what someone is entitled to do” are operationalized through identifiers and credentials. Those claims are then evaluated through verification processes that increasingly rely on inference. As these processes scale, the role of governance becomes inseparable from architecture. Trust becomes something designed, measured, audited, and continuously maintained, rather than assumed.

2. Four perspectives on the new identity ecology

Online identity can be understood through four complementary perspectives that together form a practical lens for design, deployment, and oversight: psychological, developmental, technological, and governance perspectives. Each is necessary to understand the human-AI-edge nexus.

2.1 The psychological perspective: Identity as experience

Identity is not only information; it is also experienced. Interfaces, cues, and feedback mechanisms shape self-perception, attention, and decision-making, sometimes in subtle ways. When identity-related interactions are repeatedly mediated by persuasive or predictive systems, the risk is not limited to technical error. There can be cumulative effects on autonomy, emotional equilibrium, and well-being. This is particularly salient in environments where identity is continuously inferred and where signals such as engagement, sentiment, or behavioral traces become proxies for trustworthiness. The emergence of AI-enabled deception and synthetic realities further complicates the psychological boundary between authentic and constructed identity experiences.

2.2 The developmental perspective: Congruence and well-being

Identity formation is a developmental process, and digital contexts increasingly shape how individuals experiment with self-presentation, social feedback, and belonging. In this setting, congruence between online and offline selves can be understood as a protective factor that supports coherence and resilience. Humanistic traditions emphasize the importance of congruence for healthy self-development [3]. Digital environments can either support reflective expression or reward performative identities that amplify dissonance and fragmentation. For educators and designers, the implication is practical: Identity systems and platforms should reduce coercive pressures toward self-curation and should strengthen safe mechanisms for authentic participation and redress when harms occur.

2.3 The technological perspective: Identity infrastructure at the edge

Technological change has shifted where identity processing occurs and what kinds of evidence are treated as authoritative. Cloud services and large-scale identity providers remain central in many ecosystems, yet intelligence is moving toward the edge through on-device biometrics, hardware-backed credentials, and continuous context-aware verification. This shift can reduce exposure of raw personal data and can enable privacy-preserving patterns, including local processing and selective disclosure. At the same time, it introduces lifecycle risks that cannot be ignored, such as model extraction, firmware tampering, and over-the-air update vulnerabilities. The result is a dual reality: Edge intelligence can not only enhance privacy and responsiveness, but can also expand the attack surface and complicate assurance.

2.4 The governance and security perspective: Trust as an engineered property

AI reshapes the security landscape of identity by enabling risk scoring, anomaly detection, and automated decision support, while also introducing new classes of vulnerabilities. Black-box inference can make it difficult to explain or contest identity outcomes, adversarial machine learning can undermine reliability, and synthetic media can erode the evidentiary value of familiar signals. Governance therefore becomes operational, not aspirational. Contemporary regulatory and standards landscapes increasingly emphasize accountability, transparency, and lifecycle risk management for high-impact systems, including identity-related AI [4–6]. In practice, this calls for security-by-design controls that span data, models, decisions, interactions, and oversight, with explicit mechanisms for auditability and human review where stakes are high.

3. A multilayered model of online identity

The conceptual diagram in **Figure 1** offers a simple way to organize these perspectives. The Psychological Layer captures self-perception and emotion. The Developmental Layer relates to authenticity and identity formation across life stages. The Technological Layer concerns AI-driven verification, credentialing, and data architectures, including edge-based processing. The Governance Layer embeds ethics, regulation, and accountability, translating principles into enforceable constraints.

Cross-cutting principles, trust, privacy, and autonomy connect all layers. Trust depends not only on security controls and reliable inference, but also on user perception and contestability. Privacy depends not only on minimization and safeguards, but also on how identity systems nudge disclosure and normalize surveillance. Autonomy depends not only on technical choices and policy constraints, but also on interface design and social incentives. Taken together, these layers describe a new identity ecology where cognition, computation, and compliance continuously interact.

4. The central tension: Automation and authenticity

AI offers efficiencies in verification, personalization, and fraud detection, yet it also challenges authenticity and agency. The convenience of intelligent systems can mask gradual shifts in control. Recommendation engines shape preferences, synthetic media blurs truth, and behavioral analytics can predict intentions before

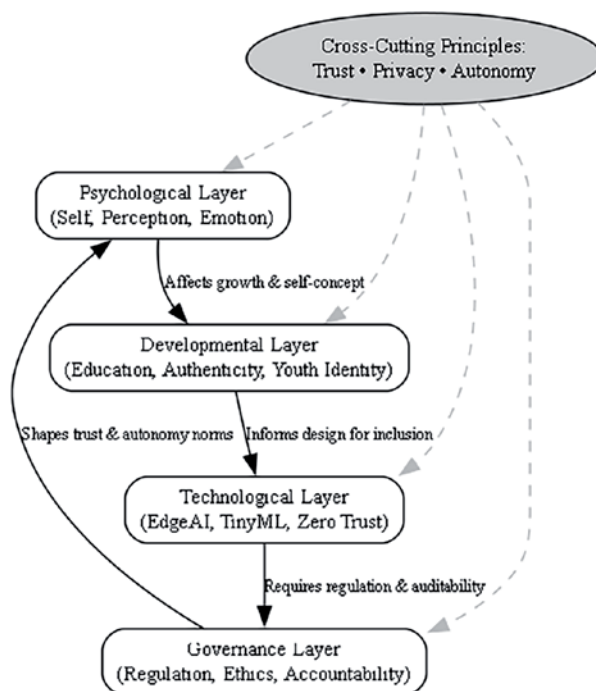


Figure 1.
Multilayered model of online identity.

they are consciously articulated [7]. The paradox is clear: More automation does not necessarily produce more autonomy, and more intelligence does not always foster more trust.

A balanced path forward requires clarity about what should be automated and what must remain contestable. Technical systems should be designed for explainability where it matters, and not as an afterthought. Privacy-preserving architectures should be adopted where feasible, and their trade-offs should be documented transparently. Educators and institutions should strengthen digital literacy so that individuals can understand how identity is inferred and how to challenge outcomes. Policymakers should ensure regulatory agility that protects rights without freezing innovation.

5. Conclusion: Toward a human-centered identity future

As physical and digital realities intertwine, the challenge is not only to engineer secure systems, but also to safeguard meaning, authenticity, and trust. Practical pathways for responsible innovation include the following:

- Integrating ethics-by-design into digital identity systems.
- Embedding explainability and accountability at every algorithmic decision point.
- Developing metrics and key performance indicators (KPIs), such as fairness scores, false-positive rates, and response latency, that translate principles into measurable governance.
- Promoting collaborative research across psychology, computer science, law, and education to ensure identity systems serve human flourishing.

The evolving landscape of online identity is not merely technological. It is civilizational. As AI systems increasingly mediate presence, perception, and participation, the task before scholars, practitioners, and institutions is clear: build identity ecosystems where technology amplifies, rather than obscures, what it means to be human.

Author details

Livinus Obiora Nweke^{1,2*†}

1 Noroff University College, Kristiansand, Norway

2 Norwegian University of Science and Technology (NTNU), Gjøvik, Norway

*Address all correspondence to: livinus.nweke@noroff.no

†These authors contributed equally.

IntechOpen

© 2026 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Sedlmeir J, Smethurst R, Rieger A, Fridgen G. Digital identities and verifiable credentials. *Business & Information Systems Engineering*. 2021;**63**(5):603-613
- [2] Nweke LO. *National Identification Systems As Enablers of Online Identity*. London, UK: IntechOpen; 2023
- [3] Rogers CR. A theory of therapy, personality, and interpersonal relationships: As developed in the client-centered framework. In: Koch S, editor. *Psychology: A Study of a Science, Formulations of the Person and the Social Context*. Vol. 3. New York: McGraw-Hill; 1959. pp. 184-256
- [4] European Union. Regulation (EU) 2016/679 of the European parliament and of the council. *Official Journal of the European Union*. 2016;**119**:1-88
- [5] European Union. Regulation (EU) 2024/1689 of the European parliament and of the council. *Official Journal of the European Union*. 2024;**1689**:1-144
- [6] NIST. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. Technical Report. Maryland, Unites States: National Institute of Standards and Technology; 2023
- [7] Williamson SM, Prybutok V. The era of artificial intelligence deception: Unraveling the complexities of false realities and emerging threats of misinformation. *Information*. 2024;**15**(6):299

Chapter 2

Dyadic Influences on the Psyche of AI Online to AI Onlife

Angelo Compierchio, Claudio Santoro and Phillip Tretten

Abstract

Emphasis on Artificial Intelligence (AI) development has too often been used as a way to justify prime instinct needs. These changes have been expressed in terms of humanized elements driven by emotional conflicts concealed unconsciously. Hence, we explore potential AI threat scenarios on subliminal persuasion breaking into the mind, constantly seeking positivity and security. This state inherent of the individual psyche raises ethical concerns about manipulation and deception, particularly in advertising and marketing. The ubiquity of AI online and its propensity to intrude on people's lives have both perpetrated profound and lasting effects on psychological and physical health. Hence, online experiences have been examined in an orderly risky give and risky take form. This designation, influencing real life decisions with the pretense of seeking harmony beyond the field of being initially focused on the self, can expand to encompass a broader perspective on reality. In practice, it can broaden people's perspectives and shape their everyday choices. In this new reality, a new self is formed, affirming all that is conceived by fostering practical onlife relations. Conversely, subliminal manipulation compels the self-attaining happiness or misery between free and constructed reality. This evidence suggests that subliminal stimuli can mask the motivational interdependency of users' behaviors with subsequent deliberation of conscious responses.

Keywords: subliminal state, neurosis, ego, superego, id

1. Introduction

Six years after Turing introduced the imitating machine, the term AI appeared. In 1990, Tim Berners-Lee followed Turing's computational communication endeavor by linking together a huge network of computers and laid instruction codes for users seeking information resources on the web. Since then, over 4 billion users rely on the internet for improving their livelihood needs [1]. Surely, the so-called mechanization (automation) that followed presupposedly maintained some acclimatized forms of equilibrium to preserve the sociological structure between people and the environment. But clearly, no man foresaw the natural adaptation of technological hypes of intelligent machines hosting far more societal interventions in grasping every form of human life. In fact, what seemed to be the most personal of our imaginative construct is now mechanized. Through the use of AI, this intervention transcends from inorganic to organic realm in establishing a symbiotic relationship with life. Consequently, the decision-making process is no longer solely personal or voluntary but is initiated

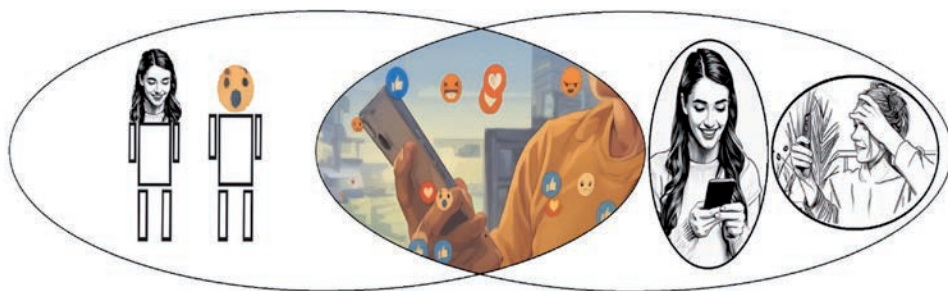


Figure 1.
Aspects of cognition and interaction with technology.

by artificial means, where motorial actions are “automatically referred to a source distinct from the self [2]”. This amalgamation influences an individual’s internal realm to the extent that we can envision the virtual disembodiment of an emoji character, transforming it from a mere symbol into a humanized representation powered by artificial intelligence (**Figure 1**).

1.1 Causes and consequences

In this chapter, we explore how the ability of users to safely manage their online personas has steadily eroded in a perceived reality. In such a space, authentic human interactions are mimicked by AI-generated contents, sparking illusory relations. We embrace the user’s interpersonal relations during social dyadic interactions by grounding machine symbols and their implications on cognition and sensory-motor events [3]. This strong human-technology bond becomes part of an evolving psychic equilibrium, influencing both humans and machines. In this unity, we reference the origin of the structure of being to Plato’s views of the indefinite dyad, also called the Great and Small, often referenced as excess and deficiency, or more and less [4].

Online platforms are being used more and more to spread hate, harassment, and abuse, which has increased by reaching a large number of people looking for harmony and well-being [5]. The rising momentum of this scenario is largely due to AI-driven threats targeting not just technical systems but also the psychological and cognitive vulnerabilities of users. This information-sharing dependency is believed to involve distinct Give and Take mechanisms without seeking a compromise and without applying conscious awareness [6]. In essence, the use of subliminal and manipulative techniques has raised concerns about autonomy, ethical boundaries, and the potential for harm through subtle influence on user behavior and decision-making. Social media “venturing into the depths of the human psyche [7]” has facilitated learning instincts appealing directly to feeling and emotions [8].

This implication indicates that an evolutionary mutation is occurring, which conceals subliminal stimuli that may bias response options. While the anthropological structures impacted are designed to strip away the characteristics of their semantic roles and sociological identities.

2. Transition from AI online to AI onlife

With online interactions, internal representational states are extended into the internet environment, while requirements for productivity are expanded from

traditional storage and processing of internal resources to external media resources [9, 10]. This favors the integral link between cognition and perception distinguished from the Jungian form of self-regulation and self-control. Whereas, users are virtually positioned between two interactive worlds, the world of being and the world of becoming. With two distinctive worlds, we conceive the great and the small paired according to Plato's metaphysical relationship of what is intelligible to one and visible for the other [11]. In this stance, we mark war-like traumatic symptoms, highlighting the "psychical conflict" in tracking a neurotic origin of personality disorders within online interactions. Generally, a neurotic symptom portrays a conflictual struggle without a valid, possible outlet, both within-person and between-person effects. This modality unearths one's personality with certain behavior and reactions of a neurotic type dominated by irrational motivations and unconscious conflicts. While, when facing debilitating external impulses, the neuronomic mode invades one's personality completely. In this case, we are faced with very serious patients, who, for example, can no longer enjoy family adaptation, professional integration, or a satisfactory relational life [12]. Alternatively, as Freud pointed out, a neurosis of a conflict or psychic trauma somehow external to the personality affects a person only in the form of a seriously disturbing event [13]. This neurosis denotes deleterious events that disturb the psychic equilibrium and destroys homeostasis. The individual dyad, in a static bond with conversation flowing by (real) humans, exchanges energetic, affective, and intellectual factors between the ego and the id containing subdued drives and unconscious defense mechanisms. This state of equilibrium is anchored in biological reality with the superego created from early childhood through the introjection of social defenses (**Figure 2**).

When interactivity guides the social discourse between the great-and-the-small bonding notion overflowing as AI-mimic humans, it generates imaginative constructs and uncanny visually guided behavior as "our perceptions are not necessarily all that we know [14]". Further, we could ask how conscious we are of our experience interlinking both images of reality and constructed reality. In sustaining this thought, we need to frame the awareness gap between image and reality and shape our perception boundaries. Without any barriers, at the outset, the growing trend of users resembled the search for affective influencing factors through means of offsetting feelings of inferiority. Within the melding of newly defined cyber-dependent categories, this drive is often reflected in users developing trauma when seeking to identify themselves through a category of a kind, "where category is a set of individuals in the environment or introspection [3]". Categorization is driven by episodic simulation, which

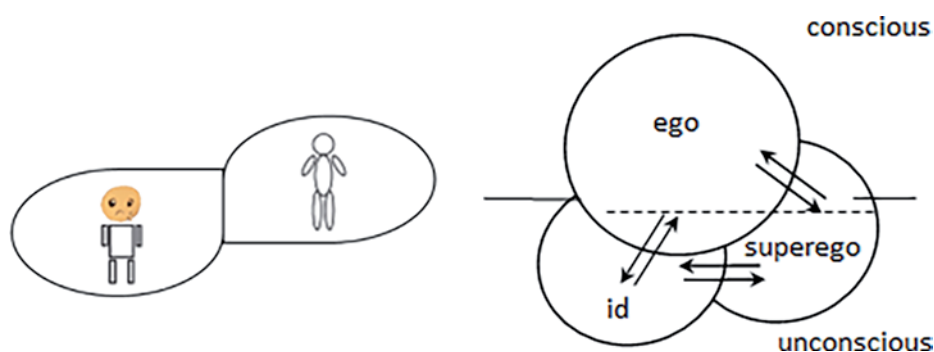


Figure 2.
 Dyadic interaction in a static bond (left) and the psychic equilibrium (right).

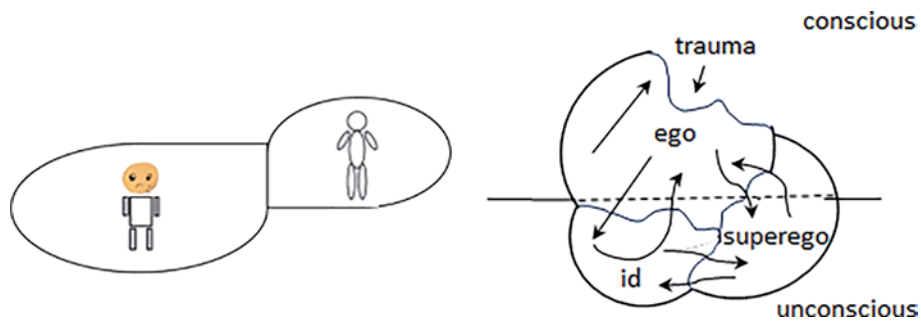


Figure 3.
Dyadic interaction in a great-and-the-small bond (left) and the unbalanced psychic equilibrium (right).

blurs the extension of one category into another without any dividing line between the real and unreal. The traumatic consequences that follows deform the ego by activating its protection mechanisms. In addition, the previously established equilibrium is thrown off balance with the risk of the id piercing the ego or the reappearance of some other impulses through casual connection. Accordingly, the implication of the great-and-the-small inquiry is inferential during preconscious processing, leaving unconscious neural representations in control (**Figure 3**).

Fundamentally, episodic memories traces are retrieved and reactivated in accordance with simulation theory, “what it is for a subject to remember ...*when the unreal becomes real*... is for him to imagine an episode belonging to his personal past. And imagining need not draw on stored information ultimately originating in experience of the relevant episode [15]”. This multi-staging exchange concatenates the ego and the superego transformation from a sense of well-being to misery, escalating aggression, anxiety, and conventional anxious dreams. The irruption of the instinctual system acts as a compensator of neurotic mechanisms, and if they are temporary, they contribute to re-establishing equilibrium. When they last too long, they turn pathological, which in certain individuals is a predisposition to the form of latent neurosis [16]. Studies carried out on people who have shown a neurotic decompensation during war or certain traumas tend to demonstrate the extent of their fragility due to psycho-affective disturbances that occurred in childhood or evolving at a later age [17, 18].

Under these conditions, personality elements are continuously changing and concurring with real-life situations. Moreover, the rise of information-sharing platforms has permanently bound human interaction through masked interactive forms. The rise of AI and its intersection with spiritual frames has transformed citizens of any social and cultural background into a community seeking spiritual wellness. In this capacity, a sense of spiritual magnetism, co-operating with our inner unconsciousness, has radiated crowded interface platforms [19].

Naturally, every form of information technology or social media platform has been built to address human genetic or environmental adaptability. Therefore, the complete avatar disguise, instinctively accepted without any conscious judgment, has permanently transformed every human needs, desires, and expectations.

3. Changes in psychic equilibrium

The spanning of the aforementioned AI-enhanced social interactionism has given users a virtual stage on which global influences are grounded on better-than-reality

perspectives. Collecting data on users' perceptual representations enabled the creation of algorithms that reveal each user's mental map. Controlling the online experience in this way depends on the strategic use of icons and symbols. They then shape the online environment through carefully crafted icons and symbols. Such a formulation has created a far-reaching analogy with Jung's symbolism, as the symbolic icons construct is presumed to carry a person "fringe of consciousness [19]". In any artificial environment, the human-AI coupling relationship parsimoniously provides a conceptual notion of intentionality founded on: (i) directiveness and (ii) condition of satisfaction [20, 21]. In this perspective, the notion of intentionality brings in a social connection that can be further represented into the engagement and disengagement of the human mind, in essence conceived and confronted with observable behavioral changes. This external relation evolves the user's embodied intentions as soon as the online display is turned on. The knowledge gained from online experience occurs when a perceptual state is formed through configured neurons being activated when representing the shape of the image. This knowledge is not peripheral to the target user as it embraces a far deep neural insight influenced by emotional signs.

3.1 Dyadic disruption mechanisms

Notably, the interpreting mind establishes a dynamic connection with the online information flow between the human and the system, starting from ordered signs labeled as Icon, Index, and Symbols [22]. The mental sensation that follows the Icon to the Index and the Symbol-mind connection selected by users raises the web as a cosmic Leviathan's crowned entity. This interaction from a philosophical perspective continues with a quest for the immortal God, releasing trust, peace, and defensiveness [23]. The first element under scrutiny is the environment, actually, the many overlayed environments constantly evolving in both space and time and remarkably set to precipitate into pre-constructed myths and symbols. However, the inescapable confrontation between two interactive users once engaged seems to move forward undisturbed, just like a moral precept. In explaining the main points of the empirical evidence collected, we need to dive into risky decision-making behaviors that often lie at the root of fraudulent communications. Although the chain of decisions that follows each interaction is closely associated with the user's personality as well as time pressures, the user is consciously relying on Jung's natural distinguishing of the inner authentic self and the triggered ego [24]. This conception addresses, in an interactive form, how information properties are shared within oneself and how they are seen and interpreted by others. In this context, the dyadic interaction represents a constructed model of evolution consisting of an adaptive fit between the organism and the world [25]. This experience is never singular or isolated but embedded within an artificial environment. The depicted psyche evolution model out of the dyad takes into consideration the environment as a mixture of probabilities of the "Give and Take" risk-taking attitudes transmitted for perception outputs between the user and the perpetrator [26]. The continued interaction over time (t) of the user (U1, U2, U3...) is alternated by perceived cyber threat evolutions of the perpetrator (P1, P2, P3...) into a Give and Take form within the altered environments (E1, E2, E3...) (**Figure 4a**).

Although the real personality of the user forms the basis of the interaction experience, when connected online, this form is highly hidden in the environmental shadow and rendered unconscious. Thus, the single environment, E1, has to be confronted in a superposition state of the Give and Take equation itself (**Figure 4b**).

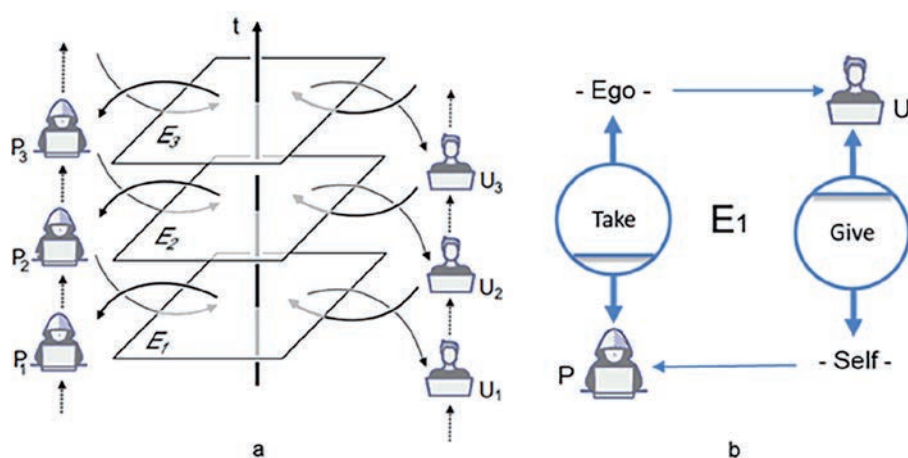


Figure 4.
A dyadic psyche evolution model (a), the Self-Give and the Ego-Take in a single E_1 environment (b).

The superposition state induces a neural orientation with continual learning utilizing biased interaction patterns while enabling new forms of experience [27]. This conjunction plays a special part in the way coupling is established for the purpose of survival [28]. The attacker itself feeds an ongoing spiral process in constructing embodied emotions, seeking the user's bodily reaction to them. Under these conditions, even a modest reply to the online attacker made the attack economically worthwhile [29].

4. Transcripts of AI onlife

The impetuous AI-based mechanization of civilization, revolved around “the only game in town [30]”, turned the socio-cultural context into a ghostly presence. The progress of AI that followed rather than enhance human potentiality was diverted towards Turing and Simon's roots “machine that think, that learn and that create [31]” all that remained was as having left the actual user behind. This emphasis shuns the contemporary Darwin in mind evolutionary quest for embryonic best minds in society. This approach in per se has revolutionized the human cognitive nexus where “cognition refers not only to universal patterns of information transformation that transpire inside individuals but also to transformations, the forms and functions of which are shared among individuals, social institutions, and historically accumulated artifacts [32]”. Nowadays, shared digital information patterns have taken on a holographic form, yet remain effectively one-dimensional. This development allows users to perceive humanized computational forms having personality features like the user themselves [33]. In every engagement, the user perceives aspects of the environment with a ready-to-hand attitude within a non-perceptual structure [34].

Personality attitudes are highly relevant within the risky cyber dimension and cannot be overlooked. In this view, both the affective qualities of anxiety and aggression of either individual in the dyad are viewed as a single personality syndrome. Traditionally, aggressiveness is seemingly discharged to cause emotional harm, while anxiety is associated with the non-discharged aggressive state. In Freudian theory, aggression is placed between two instincts: self-protection and self-destruction, with both preserving some form of aggression within a non-casual structure.

4.1 Risky take interactions

Dyadic risk-takers exhibiting aggressive, self-destructive behavior can be divided into two types: those who actively seek risk and those who disregard it. A series of Cognitive Absorption (CA) events is reported to highlight new techniques of emancipating hysterical online behaviors. CA is driven by motivational factors of an intrinsic form, exulting personal satisfaction and pleasure in ordaining behavioral expectation rewards [35]. The psychological consequences of such a traumatic exposure seem to forcibly target users' objective stimuli while evoking a sense of superiority. In fact, attackers continue to flourish and overpower online masses with elaborated messages of an emotional nature. These recurring events driven by hypnotic conditioning are mechanically driven to siphon private data from the winning candidate (Figure 5).

The receiving candidate in each sample is actively and egotistically persuaded to open the message that is offering him prizes or money subjected to tight time constraints while bypassing focused attention and rationality. As a result, the outcome of such demands prevails over cognitive control and intentional behavior. Every emailed occurrence follows a concise effort to attract the user's subliminal stimulation in storing unconsciously pieces of messages into long-term memories. As it happens, the semantic integration of novel information shapes subsequent consciously based decisions [36].

4.2 Risky give interactions

Interactions focused on giving provide a chance to acknowledge feelings and moral resilience when facing adversity. This kind of reaction is therefore subjected to critical self-reflections with a link between physical events and real life. The requests in the email samples 1–3, below, are not consciously discarded from persevering emotional responses. As a matter of the feeling involved, unreasonable bounds are bypassed and strongly felt during phased adjustments and executions in generating a response [34]. This deployment depends on the susceptibility of the receiver's thinking and decision-making process:

☆ Jurassic_PARK!	You have just RECEIVED a Free Bonus!12 - Check your invitation	Sep 3
☆ GOLDEN-TIGER	CongrATULatIONS:-YOU-...COUID-...-WIN! - Rec:-*2nd-...-ATTEMpt-...	Sep 2
☆ Attention_Required	You HAVE JUST Received a free - Millionaire PASS!	Sep 2
☆ CasinoRewards	delivery notificationCongrats: Open immediately!	Sep 1
☆ IMPORTANT!	Your- 'NZ\$1500'-ticket- 'Bonus'-inside!! 85zd94 - ebd **Congratulations**You...Could_Win!-OPEN-...	Sep 1
☆ Golden Tiger	2nd attempt Your NZ\$1500 ticket Bonus Inside!	Sep 1
☆ *OvernightMillional.	NZ\$ 3,466,6 02.99Opp orTu Nity.: 4000 - YourEmail-HAS BeEN Selected!	Aug 31
☆ BonusCodes_	Limited Time Only- A bonus Link inside!!! - Important Update - MU ST SEE!	Aug 31
☆ Zodiac Winner	Congrats Open immediately! - URGENT New Mega MOOLAH winner! I - Open NOW!	Aug 31
☆ YOUre_approved	Attention Player Your welcome Pass arrived!PV - Check your invitation	Aug 30
☆ Zodiac Winner	Congrats Open immediately! - URGENT New Mega MOOLAH winner! I - Open NOW!	Aug 30
☆ 123cards	Share your eCards! - At 123cards, we never get tired of making it as easy as possible for you to reach out to family or friends on ...	Aug 30
☆ ✦CONGRATULATIONS✦	\$5,913,537.72..Winner...Annoucement [985] - Congratulations WIN \$5913537.72 INSTANTLY!	Aug 29
☆ CAPTain-CoOKs NOTIF.	CONFIRMED - Y'OUvE BEen aWArded a Golden Ticket	Aug 29
☆ Golden_Tiger ALERT	We_HAVE Been TRYing TO_Reach--YOU D4 - NZ\$1500 Welcome BONUS	Aug 29
☆ List_-OFwinNers!	1SuRPRISe2AwAITsyou-	Aug 28
☆ --Cong'ratulation/s.	...\$1,413,537.73Win/nerAnno'uncem/ent... - ***	Aug 28

Figure 5.
Cognitive absorption of extrinsic motivational beliefs.

Sample 1:

From: John Kim <email>

Sent: XXX_X:X_AM

To: Student <email>

Subject: Re: Apple cards

I cannot handle anything here at the hospital that is why I said you should get me the Italian Apple iTunes gift card for me tomorrow if you got money from your parents.

Sent from my iPhone

Sample 2:

From: John Kim <email>

Sent: XXX_X:X_AM

To: Student <email>

Subject: Re: Apple cards

Hello Student,

How are you doing this morning? I have been trying to get the Paygoo prepaid card but it's not working and I do not why, they did not accept my info. Please kindly try your possible best and get me just 100€ Italian Apple iTunes gift card and I promise you once I'm out from here am going to send you all your that you have spend in buying does cards for you and I will also reimburse you then add compensation for you,

Sent from my iPhone

Sample 3:

From: John Kim <email>

Sent: XXX_X:X_PM

To: Student <email>

Subject: Re: Apple cards

I cannot handle anything here at the hospital that is why I said you should get me the Italian Apple iTunes gift card for me tomorrow if you got money from your parents

Sent from my iPhone

Users reflexively undo difficult choices when they detect a moderately risky shift during give interactions. The level of persuasive influence is not authoritarian; rather, it is conducted on a more personal level with the aim of discontinuing any suspicious activity.

5. Conclusion

The current AI direction has been set to go beyond Turing and Simon's expectations, with a driven dehumanization increasingly repellent to the earth's natural environment. In addition, this rising trend delimits humans' decisions in whether to accept or otherwise AI tools in their working and social domains. In all this regard, there is a continuous unwaning of risky Give and Take interactions aimed at taking away distinct cognitive processes in making up critical thinking, creativity, and dignity, to name a few, in exchange for a little comfort or immediate profits.

Given that AI-driven threats exploit users' psychological and cognitive vulnerabilities, there is growing concern that such attacks could undermine humans' inalienable rights. These rights include the right to life, freedom, and liberty, which are not transferable and cannot be taken away. But inalienable rights do not exist unless they are enforced. If being human means creating something abstruse and completely different from oneself, then we stipulate that one tries to get closer to God. This emphasis leads to closely examine the impacts of AI on the psyche of the masses, and humanity itself. In 1983, during the height of the Cold War, Colonel Petrov witnessed a false nuclear alarm, he disobeyed orders and saved the almighty world! Would had been an AI agent in his place, the outcome would have been indescribable. Good Petrov asked himself, it is not possible that they have unleashed this war tonight. There was practically a magnetic earthquake on his monitor that made him glimpse these missiles, which were not missiles. But it was his humanity that at a certain point opposed the screen machine, and he did not raise the alarm to obviously launch the missiles from Russia toward America. The key question is simple: what did humanity learn from this experience?

That it rests with humanity governing this new reality, like all the other new realities that can emerge.

Conflict of interest

The authors declare no conflict of interest.

Author details

Angelo Compierchio^{1*}, Claudio Santoro² and Phillip Tretten¹

1 Luleå University of Technology, Luleå, Sweden

2 OsteoMedica, Benevento, Italy

*Address all correspondence to: angelo.compierchio@associated.ltu.se

IntechOpen

© 2025 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Bailey J. Communication: Telephone, computers and WWW. In: *Inventive Geniuses Who Changed the World: Fifty-Three Great British Scientists and Engineers and Five Centuries of Innovation*. Cham, Switzerland: Springer International Publishing; 25 Nov 2021. pp. 363-401. DOI: 10.1007/978-3-030-81381-9_15
- [2] Jeannerod M. Consciousness of action. In: *The Blackwell Companion to Consciousness*. India: John Wiley & Sons Ltd.; 12 Apr 2017. pp. 576-587
- [3] Barsalou LW. Perceptual symbol systems. *Behavioral and Brain Sciences*. 1999;22(4):577-660
- [4] Nyvlt MJ. Aristotle and Plotinus on the Intellect: Monism and Dualism Revisited. UK: Lexington Books; 2012
- [5] Guembe B, Azeta A, Misra S, Osamor VC, Fernandez-Sanz L, Pospelova V. The emerging threat of ai-driven cyber attacks: A review. *Applied Artificial Intelligence*. 2022;36(1):2037254. DOI: 10.1080/08839514.2022.2037254
- [6] Sherry DF, Schacter DL. The evolution of multiple memory systems. *Psychological Review*. 1987;94(4):439. DOI: 10.1037/0033-295X.94.4.439
- [7] Al-Hashem N, Saidi A. The psychological aspect of cybersecurity: Understanding cyber threat perception and decision-making. *International Journal of Applied Machine Learning and Computational Intelligence*. 2023;13(8):11-22
- [8] Marler P. The instinct for vocal learning: Songbirds. In: *Plasticity of Development*. Cambridge, MA, USA: MIT Press; 15 Jul 1991. pp. 107-125
- [9] Donald M. *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge, MA, USA: Harvard University Press; 1993
- [10] Wells AJ. Turing's analysis of computation and theories of cognitive architecture. *Cognitive Science*. 1998;22(3):269-294
- [11] Plato. *Plato in Twelve Volumes*, Vols. 5 & 6 translated by Paul Shorey. Cambridge, MA; London: Harvard University Press; William Heinemann Ltd; 1969. [Internet]. 1987. Available from: <https://www.perseus.tufts.edu/hopper/text?doc=urn:cts:greekLit:tlg0059.tlg030.perseus-eng1:7.524c>
- [12] Felluga D. Modules on freud: On neuroses. In: *Introductory Guide to Critical Theory* [Internet]. 2011. Available from: <http://www.purdue.edu/guidetotheory/psychoanalysis/freud4.html>
- [13] Freud S, Jones E, Abraham K, Ferenczi S, Simmel E. *Psycho-Analysis and the War Neuroses*. No. 2. Vienna, Austria: International Psychoanalytical Press. EBook #55443, DigiCat; [Internet] 2022. Available from: https://mindsplain.com/wp-content/uploads/2020/08/Psycho-analysis-and-the-War-Neuroses-by-S.-Ferenczi-Karl-Abraham-Ernst-Simmel-and-Ernest-Jones_.pdf
- [14] Reams J. Consciousness and identity: Who do we think we are? *New Ideas in Psychology*. 1999;17(3):309-320
- [15] Michaelian K. *Mental Time Travel: Episodic Memory and our Knowledge of the Personal Past*. Cambridge, MA, USA: MIT Press; 2024
- [16] Inouye E. Genetic aspects of neurosis a review. *International Journal of Mental Health*. 1972;1(1-2):176-189

- [17] Cresti L, Suman A. Identity and change, chapter 1. In: Cresti L, Lapi I, editors. *Psychoanalytic Psychotherapy Between Identity and Change*. Oxon, UK, NY, USA: Taylor & Francis; Routledge Publishing; 29 Nov 2024
- [18] TenHouten WD. *Alienation and Affect*. London: Routledge; 2017
- [19] Jung CG. *Man and his Symbols*. NY, USA: Bantam; 2012
- [20] Gallagher S. Pragmatic interventions into enactive and extended conceptions of cognition. *Philosophical Issues*. 2014;24(1):110-126
- [21] Madzia R, Jung M. Introduction: What a pragmatist cognitive science is and what it should be. In: Jung M, Madzia R, editors. *Pragmatism and Embodied Cognitive Science*. De Gruyter Publishing; 2016. pp. 1-14. DOI: 10.1515/9783110480238-001
- [22] Peirce CS. *Collected Papers of Charles Sanders Peirce*. Cambridge, MA, USA: Harvard University Press; 1934
- [23] MacKinnon RC. *Searching for the Leviathan in Usenet*. USA: San Jose State University; 1992
- [24] Lauriola M, Levin IP. Personality traits and risky decision-making in a controlled experimental task: An exploratory study. *Personality and Individual Differences*. 2001;31(2):215-226. DOI: 10.1002/mar.1029
- [25] Sterelny K. Minds: Extended or scaffolded? *Phenomenology and the Cognitive Sciences*. 2010;9(4):465-481
- [26] Khrennikov A. Quantum-like model of unconscious-conscious dynamics. *Frontiers in Psychology*. 2015;6:997
- [27] Dimyan MA, Cohen LG. Neuroplasticity in the context of motor rehabilitation after stroke. *Nature Reviews Neurology*. 2011;7(2):76-85
- [28] Dewey J. *Experience and Education*. New York, NY, USA: Simon & Schuster; 1938
- [29] Norton email scams: Answers to FAQ [Internet]. 2022. Available from: <https://uk.norton.com/blog/emerging-threats/norton-email-scams>
- [30] Nersessian NJ. Interpreting scientific and engineering practices: Integrating the cognitive, social, and cultural dimensions. In: *Scientific and Technological Thinking*. UK: Psychology Press; 2004. pp. 17-56
- [31] Simon HA, Newell A. Heuristic problem solving: The next advance in operations research. *Operations Research*. 1958;6(1):1
- [32] Nersessian NJ. *The Cognitive Basis of Model-Based Reasoning in Science*. Cambridge, UK: Cambridge University Press; 2002
- [33] Cook J, Salvendy G. Perception of computer dialogue personality: An exploratory study. *International Journal of Man-Machine Studies*. 1989;31(6):717-728
- [34] Heidegger M. *Being and Time*, trans. J. Macquarrie. Robinson, Harper & Row; 1962
- [35] Tononi G. An information integration theory of consciousness. *BMC Neuroscience*. 2004;5:1-22
- [36] Agarwal R, Karahanna E. Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly*. 2000;24(4):665-694

‘Digital Identity Congruence’: Exploring the Importance of Congruent Online Identities for Adolescents

Jennifer Laffier and Madison Westley

Abstract

Social media platforms such as Instagram, TikTok, and Snapchat have become prominent arenas for adolescent identity exploration and expression. However, the fluidity and duality of online self-presentation introduce new psychological risks, including digital identity incongruence, which stems from a lack of alignment between one’s online digital identity and authentic self, leading to psychological distress. This chapter explores how digital identity congruence can play a critical role in supporting psychological well-being during this formative developmental stage. Based on theoretical and empirical research, four key areas of intervention are proposed: (1) Positive Youth Development (PYD) approaches that empower identity formation, (2) supportive relationships that foster authenticity and social acceptance, (3) promotion of self-awareness and self-acceptance as protective factors, and (4) technology design and policy focused on health promotion and congruent self-expression. These insights demonstrate that congruence can be developed through reflective digital engagement, supportive relationships with peers and adults, and intentional technology use. This study contributes to the growing fields of youth, identity, mental health, and technology, by identifying congruence as both a developmental need and protective factor for adolescent well-being. Implications for educational programming, psychological and mental health interventions and practices, and ethical technology design are also discussed.

Keywords: adolescence, identity formation, technology, psychological health, development

1. Introduction

Oscar Wilde’s oft-quoted aphorism, “Be yourself; everyone else is already taken,” [1] captures the enduring tension between individuality and conformity. This tension is particularly pronounced during adolescence, a developmental stage marked by the dual imperatives of self-definition and social belonging [2–4]. Identity formation

involves integrating self-concepts, such as personal interests, values, and gender identity, into a coherent and stable sense of self [2–4]. As adolescents re-evaluate early identifications, they often explore new peer groups, esthetics, hobbies, belief systems, and modes of self-expression [5]. While this is developmentally normative, prolonged or excessive identity uncertainty has been linked to heightened risks of anxiety, depression, and social maladjustment [4, 6–8].

These challenges are amplified by the digital landscape in which today's adolescents are immersed. Platforms like Instagram, TikTok, and Snapchat have become key arenas for identity exploration, enabling self-presentation through visual storytelling, self-disclosure, and peer interaction [9]. Instagram showcases curated grids that reflect personal esthetics and affiliations; TikTok encourages persona experimentation; Snapchat offers more intimate, ephemeral spaces for identity exploration [10, 11]. These platforms act as both mirrors, reflecting self-concepts, and canvases for constructing desired identities.

However, the fluidity of online self-presentation introduces psychological risks. Carl Rogers [12] argued that psychological well-being stems from *congruence*, the alignment between one's self-image and authentic experience [13]. In contrast, *incongruence*, or a mismatch between internal and projected selves, is associated with emotional distress, low self-esteem, and anxiety [14]. Digital Identity Congruence [15, 16] builds on Rogers' theory, applying it to the digital domain. It describes the alignment between one's online and offline selves. It recognizes that adolescence is a time of heightened incongruence due to ongoing identity formation, during which youth often experiment with idealized or socially acceptable online personas. When significant discrepancies exist between these portrayals and the authentic self, mainly if unrecognized or sustained over time, mental health risks may increase [15].

Research confirms that adolescents frequently present hyper-idealized digital selves, intensifying vulnerability to mental health challenges [17]. This is exacerbated by pressures to conform to digitally constructed beauty norms, success, and popularity, amplified by algorithmic visibility and peer validation [18, 19]. Given these dynamics, it is critical to support adolescents in cultivating digital congruence and psychological well-being.

This chapter expands upon the current literature on youth identity development to critically examine the construct of digital identity congruence and its implications for adolescent mental health. Drawing on theoretical and empirical research, we argue that digital identity congruence (i.e., when adolescents' experience alignment between their online and offline lives) is a critical but underexamined factor in their well-being. Our analysis of the literature demonstrates that incongruence can be amplified by technology design and social pressures, while congruence can be experienced through targeted psychological and educational supports. We begin by exploring adolescence from a developmental perspective, including identity formation as a core developmental task and the role of technology in this task. Then, we introduce Carl Rogers' theory of congruence and the importance of authentic self-expression, self-acceptance, and identity integration in adolescents' identity formation. Building upon Laffier's [15, 16] concept of digital identity congruence, we examine how youths' online environments can support or hinder their identity formation and overall well-being.

We conclude by providing psychological interventions and education initiatives focusing on four core themes: (1) focusing on Positive Youth Development theory through integrating practices that promote competence, confidence, connection, character, and caring; (2) supporting youth in developing skills, knowledge, and

attitudes, such as emotional intelligence, self-regulation, and digital literacy; (3) creating opportunities for positive relationships and social acceptance online and offline; and (4) rooting technology design and policy in health promotion and opportunities for congruent self-expression. These insights contribute to a growing body of literature on youth, identity, and technology by situating the role of congruence as a protective factor for youth well-being and development. These recommendations can support educators, parents, and community members as they help youth navigating adolescence and identity formation in the digital age. They can also inform curriculum development, digital literacy programming, best practices for mental health professionals, educators, and parents working with youth, and guidelines for safe and healthy youth-centered online environments and platform design.

2. Adolescent identity formation

2.1 Adolescence

Adolescence is a critical developmental period marked by rapid physical, cognitive, and psychosocial changes. The World Health Organization [20] defines adolescence as spanning ages 10 to 19, while the United Nations and UNESCO [21] extend the category of youth to ages 15 to 24, acknowledging the prolonged transition to adult autonomy, a broader age range recognized in this chapter. This stage of life is foundational for establishing behaviors and identities that shape lifelong well-being. Today's adolescents navigate a complex landscape of challenges and opportunities, including rising mental health concerns, climate anxiety, digital safety risks, and social inequities [22]. In response, the field has shifted from a deficit-focused model to a strengths-based Positive Youth Development (PYD) approach, which views adolescence as a period of opportunity and emphasizes empowerment and flourishing [23]. This has informed educational and community practices that now prioritize youth-centered, participatory strategies, recognizing adolescents as active agents in their development.

This perspective is especially important in the digital age, where youth, not technology, should drive their own developmental trajectories. Contemporary research emphasizes that adolescent development is deeply embedded in social and cultural contexts. Influences such as family, peers, media engagement, and societal norms significantly shape behavior, well-being, and identity [11, 24]. However, adolescents are not passive recipients of these forces; they actively interpret and respond to their environments. Understanding this dynamic interplay between individual agency and social context is essential to supporting healthy adolescent development today.

2.2 Identity formation as a developmental task

A central developmental task during adolescence is the formation of identity, a process involving the integration of self-understanding, values, goals, and behaviors into a coherent and stable sense of personhood [2, 25]. Constructing a meaningful identity provides resilience and promotes adaptive functioning [24, 26]. This process is shaped by both internal reflection and external social contexts, emerging through dynamic interactions with family, peers, mentors, schools, and now, increasingly, digital environments [25]. These relationships act as social scaffolds that support adolescents' exploration of values, roles, and beliefs.

Classic and contemporary theories provide frameworks for understanding adolescent identity. Erikson's psychosocial theory describes adolescence as a period of identity vs. role confusion, where individuals seek a sense of uniqueness while maintaining affiliation with others [2, 27, 28]. Marcia [29] expanded this with the identity status paradigm, which frames development through exploration and commitment [30, 31]. Marcia [29] suggested that youth need opportunities and motivation for exploration before their achievement.

Contemporary perspectives build on these foundations by emphasizing the dynamic, context-sensitive nature of identity. For example, Waterman's [32] theory of self-realization emphasizes the role of meaningful, value-aligned activities in shaping identity. Crocetti's model presents it as an ongoing process of exploration, reflection, and commitment [33]. More recent theories, such as Identity Control Theory and Identity Shift Theory, explore how identity is shaped by feedback and social performance, suggesting that individuals maintain multiple, evolving identities that respond to external cues [34, 35].

2.3 Identity formation in the digital world

Considering 95% of teens have smartphone access and half report being online "almost constantly," [36, 37] the digital world plays a central role in adolescent identity development [38, 39]. Youth engage with technologies such as gaming, AI, media, and social media to explore, experiment, and express identity, through both active use (e.g., creating content, playing games) and passive consumption (e.g., algorithm-driven content), which influence their sense of self [40]. As Waterman stated, identity formation is about creation and discovery, [41] which aligns well with the technologies available to youth.

Gaming, which 85% of youth engage with, [42] offers identity exploration through avatar creation, moral decision-making in virtual worlds, and role experimentation [43–45]. Artificial intelligence (AI) influences youth directly through chatbots and indirectly via algorithm-driven media [16]. Conversations with AI may scaffold identity development by offering immediate, personalized feedback [46]. Media such as movies, books, music, and podcasts help youth understand themselves and their place in the world [47, 48].

Social media is highly influential on adolescent identity formation, with many spending an average of 4.8 hours daily on platforms like TikTok, Instagram, and Snapchat [37, 49]. These platforms support identity development through self-presentation, social comparison, and peer interaction [38, 50]. For marginalized youth, they offer affirming spaces [51, 52]. Yet, curated content, such as selfies taken by over 82% of teens, [53] can promote unrealistic self-images. Negative feedback, cyberbullying, and constant comparison contribute to anxiety, low self-esteem, and identity-related stress [54–56].

3. The need for congruence in identity development

3.1 Introduction to congruence

Research on identity formation emphasizes the importance of authentic self-representation, self-acceptance, and identity integration for healthy development [38, 50, 57, 58]. A coherent, consistent sense of self is linked to higher self-esteem,

psychological well-being, and positive adjustment, [59, 60] whereas fragmented or incoherent self-presentations, especially when prolonged or lacking awareness, can hinder identity development [17, 61].

The concepts of integration and self-acceptance were central to psychologist Carl Rogers' theory of congruence, which is viewed through a therapeutic lens. Rogers defined congruence as "the accurate matching of experiencing and awareness," ([62], p. 339) emphasizing its centrality to psychological well-being and personal growth. In contrast, incongruence arises when there is a discrepancy between "the actual experience of the organism and the self-picture of the individual insofar as it represents that experience" ([12], p. 213). He argued that psychological tension arises when individuals deny significant sensory and emotional experiences, thereby preventing these from being symbolized and integrated into their self-concept ([12], p. 510). Rogers believed individuals are inherently motivated to feel, behave, and experience in ways that are consistent with their self-image and reflective of their ideal self [62].

Rogers' work was elaborated on to emphasize that congruence can be achieved when an individual's perceived self (self-image) aligns with who they are and their ideal self (who they aspire to be), contributing to psychological well-being and a stable sense of identity [13]. Alignment occurs from greater self-awareness, self-acceptance, and openness to internal and external experiences. The closer the alignment between the self-image and the ideal self, the greater the level of congruence and the stronger the individual's sense of self-worth. Incongruence happens when our perceived self, authentic self, and ideal self do not match, leading to poor mental health [13]. Rogers proposed that within a trusting, non-judgmental relationship that facilitates self-exploration, individuals can become more open to their experiences and true selves, and are less defensive [63]. Rogers [62] posited that congruence is a necessary condition for achieving self-actualization, or the realization of one's full potential.

Congruence is supported by related theories examining internal conflict. *Ambivalence* [64, 65] refers to simultaneous contradictory feelings or attitudes, which can create uncertainty, as seen in suicidal ideation [66, 67]. *Cognitive dissonance* [68] describes the discomfort from conflict between beliefs and behaviors, prompting individuals to adjust either to restore consistency [69, 70]. For example, a youth may bully someone they like due to peer pressure, then justify it to reduce internal conflict [69]. Cornelius-White [71] proposed a five-dimensional model of congruent functioning: flow, genuineness, symbolization, authenticity, and organismic integration. This model presents congruence as a dynamic, ongoing process grounded in self-trust, awareness, and openness. Central to this process is *self-acceptance*, also known as unconditional positive self-regard, [72] which fosters personal integration and growth by promoting deeper relational and emotional awareness.

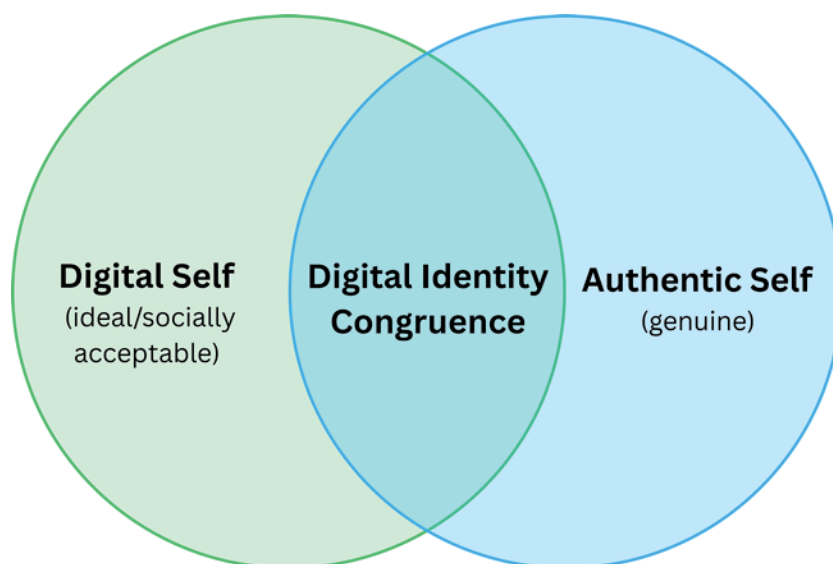
3.2 Digital identity congruence

Laffier [15, 16] expanded Rogers' theory of congruence to the digital realm, introducing the concept of digital identity congruence. This model examines how individuals, particularly adolescents, present aspects of their self-identity in digital spaces (e.g., social media, chat rooms, and gaming platforms) and through digital media (e.g., A.I. or V.R.). Drawing on research in positive youth development, identity formation, and psychological well-being in the digital age, the model posits that mental health is supported when one's online self-presentation aligns with their authentic self, an honest recognition and acceptance of one's strengths and inherent traits (**Figure 1**).

Congruence occurs when the ideal self, often an aspirational or socially desirable digital identity, aligns with the authentic self, characterized by self-awareness, acceptance, and genuineness [15]. Authentic self-representation and identity coherence are critical to psychological well-being and self-esteem [50, 57–59]. Conversely, a fragmented or incongruent digital self, where one's online persona deviates significantly from reality, has been linked to psychological distress, including anxiety and depression [17, 61].

Adolescence is a normative period of identity exploration, [24, 26] and some deviation from the authentic self may support self-discovery [27, 29]. However, excessively distorted or unrealistic digital portrayals may compromise well-being [15, 73]. As Identity Shift Theory [35] suggests, adolescents may adjust their offline identity to reflect their online persona. While beneficial when aligned with personal strengths, such shifts can be detrimental when based on inauthentic or harmful self-representations.

Essential to digital identity congruence are authenticity, self-awareness, and self-acceptance. Authenticity is foundational to identity congruence and fostering trust and meaningful relationships in adolescence and beyond. Social Penetration Theory [74] supports this, positing that interpersonal relationships develop from initial, surface-level interactions to deeper, more intimate connections through a gradual process of self-disclosure and authenticity [74]. However, the adolescent years bring challenges for being authentic in online spaces, and this is where extra guidance is needed. Self-awareness plays a vital role in promoting congruence. As adolescents engage in various forms of digital self-presentation, the ability to critically reflect



1. Alignment promotes congruence and psychological health.
2. Self-awareness is key for digital identity congruence.
3. Adolescence is a period of incongruence; problematic when excessive, prolonged or without awareness.
4. Self-acceptance promotes authentic identity formation.

Figure 1.
Digital Identity Congruence (from Laffier, 2016) [15].

on motivations and consequences becomes crucial. This reflective capacity enables recognition of inconsistencies between online and offline selves, fostering a more coherent identity [75]. Self-awareness encourages intentional, authentic online behavior that aligns with one's evolving self-concept.

Equally important is the role of self-acceptance. As adolescents cultivate self-acceptance, embracing both their strengths and imperfections, they become more capable of expressing an authentic digital self. This authenticity, in turn, supports greater congruence between their internal identity and their public digital persona, contributing positively to psychological well-being [57]. Self-acceptance is not a static trait but rather a developmental process characterized by growth, reflection, and the cultivation of unconditional positive regard for oneself. Rogers [12] explained that congruence and psychological health are fostered when individuals accept themselves fully and experience alignment between their self-concept and lived experience. Recent research affirms that self-acceptance is linked to agency and resilience in digital contexts [13, 76].

With appropriate education, guidance, and opportunities for reflection, adolescents can develop the skills needed to foster congruence. This process empowers youth to be active agents in their identity formation. While parents, educators, and digital technologies provide necessary scaffolding, adolescents ultimately steer their development through intentional, reflective, and authentic self-construction.

3.3 The need for digital identity congruence

As digital technologies become increasingly embedded in daily life, adolescents' identity formation has expanded beyond offline contexts into virtual spaces. In these online environments, youth frequently engage in curated self-presentation, emphasizing idealized traits to gain peer approval and social validation [77, 78].

3.3.1 Strategic self-presentations

Active social media engagement, such as creating and sharing content, is significantly linked to identity exploration in adolescents [77] and also concerns digital identity incongruence. Adolescents often engage in strategic self-presentation, including ingratiation, in which individuals aim to appear likable and self-promoting, seeking social approval by highlighting their achievements or competencies [78]. While such behaviors can support identity development when based on genuine traits, they may also foster identity distortion when exaggerated or fabricated. Youth may selectively share images or narratives on social media that depict an affluent, adventurous, socially desirable, or effortlessly happy life, even when these portrayals are misaligned with their actual lived experiences [79]. This discrepancy between one's authentic self and digital persona can lead to digital identity incongruence, resulting in increased psychological strain, including heightened stress, anxiety, and diminished self-worth, particularly when online validation becomes central to one's self-esteem [77].

3.3.2 Relationships and audience influence

Strong relationships develop from reciprocal and authentic self-disclosure, deepening over time [74]. However, social media often provides what has been termed "empty calories of social interaction," because interactions often lack substance and

exacerbate adolescents' heightened sensitivity and internalization of social feedback [80]. Audiences on social media, including acquaintances and strangers, can influence how adolescents perceive themselves (self-concept validation), particularly when the feedback received is negative [55]. In addition, social comparison with peers has been found to increase psychological stress and symptoms of anxiety, contribute to low self-esteem as the individual feels less than their peers, or that they cannot keep up with social trends [56].

3.3.3 Authenticity

Conversely, authentic self-expression online, the ability to disclose both strengths and vulnerabilities, is linked to greater coherence between digital and offline identities, as well as improved psychological well-being [57, 81]. Bailey et al. [57] found that active use of social media (e.g., posting, communicating with other users) contributed to positive well-being, whereas passive consumption (e.g., scrolling, reading others' posts) was negatively related to well-being. They found that authentic self-expression on social media was correlated with greater life satisfaction and that the notion of posting authentically was associated with more positive moods [57].

Recent social media movements encouraging authenticity (e.g., unfiltered posts, sharing struggles) exemplify how “being real” fosters self-acceptance and connection. Research on the social media app “BeReal” found that youth appreciated the space because it allows authentic self-presentation, where users share their unfiltered lives and engage in relatively little social comparison [80]. However, achieving digital identity congruence can be challenging for adolescents due to social pressures. Expressing authentic thoughts or feelings online, which sometimes include vulnerability, can expose youth to cyberbullying or social rejection, complicating efforts to maintain authenticity online for adolescents [54].

3.3.4 Design architecture

The design architecture of social media platforms can make congruence difficult for youth. Social media platforms leverage algorithms that often amplify idealized or performative content, subtly incentivizing adolescents to prioritize external approval and social validation over internal self-acceptance. As a result, youth may feel compelled to present exaggerated or filtered versions of themselves in order to gain likes, followers, and positive feedback, reinforcing a cycle of identity distortion and diminishing opportunities for authentic self-expression [9]. This dynamic complicates adolescents' efforts to develop a coherent sense of self and may contribute to increased digital identity incongruence.

3.3.5 Merging of online and offline lives

The digital context further blurs the boundaries between online and offline life. Mace [82] argues that individuals increasingly live merged online and offline existences. Adolescents often struggle to distinguish between their online and offline identities, leading to a complex and confusing presentation of the two [82]. For example, while some youth design avatars to mirror their authentic selves, many adopt idealized personas that represent enhanced versions of who they wish to be [43]. Although this can support developmental exploration and creativity, prolonged immersion in such idealized identities may contribute to self-concept distortion or

reduced self-esteem [83, 84]. They can also experience a distorted sense of reality when prioritizing gameplay over real-world interactions, consequently diminishing their social awareness and well-being [85]. Adolescents have even indicated that “their actual world seems impoverished compared with the opportunities and excitement available to them in the virtual worlds” ([86], p. 14). Digital identity incongruence forms as the adolescent becomes not only more immersed, but also more satisfied with their ‘online persona’, weakening self-acceptance of their authentic self.

These dynamics in youths’ digital worlds underscore the transactional and risky nature of digital self-presentation, wherein adolescents continuously negotiate identity through audience feedback and contextual demands [78]. Over time, youth who habitually perform identities divergent from their real selves may struggle with authenticity, leading to self-silencing and problematic internet use [87, 88]. While persistent identity incongruence has been linked to elevated risks of social maladjustment and mental health problems, [89–91] fostering digital identity congruence can support psychological resilience and identity coherence [15, 57, 89].

The Digital Identity Congruence model acknowledges that to promote congruence, individuals, especially adolescents, need to be equipped with the self-awareness and critical thinking skills necessary to navigate online spaces while staying grounded in their authentic selves. Acknowledging both the opportunities and risks of digital self-expression is crucial for fostering healthy identity development in today’s increasingly connected world. Additionally, individuals need to develop self-acceptance, as it facilitates the development of congruence.

4. Promoting digital identity congruence

To support adolescents through the identity formation stage, a positive youth development (PYD) lens can guide efforts to promote digital identity congruence. While online engagement presents risks, it also offers opportunities for youth to develop resilience and an authentic, self-accepting identity. Rather than approaching digital exploration with fear or restriction, adults, including parents, educators, counselors, and tech developers, can play a supportive role by fostering skills, awareness, and environments that encourage congruence. Youth should be viewed as active agents in their own development. Strategies to promote digital identity congruence are outlined across four key areas.

4.1 A focus on positive youth development (PYD)

A key principle of Positive Youth Development (PYD) is viewing adolescence as a time of opportunity rather than vulnerability, fostering empowerment and flourishing [23]. Identity formation, from this perspective, is an opportunity for adolescents to realize their potential.

Encouraging adolescents to identify and explore their strengths through creative, playful activities supports healthy identity development and digital identity congruence. Parents, educators, and community members can facilitate activities and discussions—both online and offline—that promote self-awareness and recognition of personal strengths. Adults can create opportunities for youth to develop the five core competencies of PYD: competence, confidence, connection, character, and caring. Self-reflection exercises (e.g., journaling, classroom activities) can help adolescents articulate their values and strengths, enhancing confidence. Technologies such as

social media, gaming, and communication apps can foster meaningful connections across digital and real-life contexts. When the five C's are supported, a sixth, 'contribution' emerges, reflecting a youth's ability to impact their community positively [92].

4.2 Development of skills, knowledge, and attitudes

Research highlights key moderating factors that support healthy identity formation in the digital age. Skills such as self-reflection, emotional intelligence, self-regulation, and digital literacy help mitigate risks linked to digital identity incongruence [15, 93]. Providing adolescents with opportunities to build and apply these skills is essential during this formative stage.

4.2.1 Identity-focused interventions

Identity-focused programs, grounded in PYD, are evidence-based programs that support adaptive identity formation through skill-building, reflection, and supportive environments [25]. These interventions reframe adolescence as a period of opportunity and empowerment, rather than risk, promoting self-exploration and identity congruence. Identity interventions can be self-guided (e.g., using journals, apps, or books) or collaborative, using participatory approaches that involve peer co-learning through reflection and dialog. A central component across all formats is structured reflection, which fosters self-awareness. Soh et al. [61] found that in-person group discussions support broad identity exploration by offering diverse perspectives, while solitary reflection, often supported by social media and digital resources, can help youth explore and commit to specific goals. Digital tools, such as social media and gaming, can enable connection and creative expression. For example, adolescents may use avatars, word clouds, digital art, or photo albums to explore and express their evolving sense of self. Such programs have been successfully implemented in classrooms and therapeutic settings [94, 95].

4.2.2 A sense of empowerment

A positive self-identity is closely linked to empowerment, characterized by a sense of control, influence, self-efficacy, access to resources, and skill development [15, 16, 96]. This relationship is cyclical: as youth feel more empowered, their self-esteem and self-image improve, further reinforcing their sense of agency. Youth can foster self-empowerment by using their voices to drive positive change and participating in social justice efforts. Online activism provides a meaningful way to express identity and advocate for causes that are connected to one's lived experiences, such as gender, mental health, or youth homelessness. These forms of engagement support digital identity congruence by aligning self-expression with authentic values and social concerns. Participation in community and advocacy efforts strengthens collective identity as well [10, 16].

4.2.3 Building digital literacy

As adolescents navigate increasingly digitized social environments, they must develop a nuanced understanding of the role technology plays in shaping their identity and mental health. This is particularly important given the growing pressures to conform to algorithmically reinforced norms of beauty, popularity, and success, norms that are often amplified by mechanisms of algorithmic visibility and peer feedback [18, 19]. A comprehensive digital literacy framework should include instruction

on the psychological and social dynamics of peer validation, such as the impact of likes, shares, and comments, which can significantly influence adolescents' self-perceptions and online behavior. Youth should also be challenged to critically examine how they curate and present their digital identities, including the development of hyper-idealized or selectively authentic selves. Digital literacy education should help youth recognize the value of internal validation and cultivate self-worth that is not contingent on external feedback [96, 97]. By encouraging reflective self-awareness, digital literacy education can empower adolescents to engage with digital spaces in more authentic and psychologically resilient ways.

4.2.4 Developing emotional intelligence

Educators, caregivers, and community stakeholders play a pivotal role in supporting adolescents' development of emotional intelligence (EI) and social-emotional learning (SEL), both of which are foundational for navigating digital spaces and constructing a coherent sense of identity. SEL frameworks encompass a range of competencies, including self-awareness, self-regulation, social awareness, relationship skills, empathy, and responsible decision-making [98]. Self-awareness, a core component of both EI and SEL, equips youth with the ability to recognize and understand their own emotions, values, beliefs, and behavioral patterns. This reflective capacity allows them to assess whether their online self-presentations align with their authentic selves. As youth become more attuned to internal inconsistencies, they are better positioned to make intentional decisions about how they engage in digital spaces. For instance, a common objective in adolescent social media use is to project an image of effortless happiness, health, and success, [79] masking the complexity of life and emotional experiences. Therefore, emotional intelligence education could help normalize the full range of human emotions and enable youth to recognize that psychological well-being is not synonymous with perpetual happiness.

However, there exists a significant dilemma: while authenticity and vulnerability are essential to psychological health, they can increase adolescents' exposure to negative social feedback, including cyberbullying, ridicule, or rejection [54]. As such, fostering SEL skills must be accompanied by practical strategies for digital boundary-setting and self-protection. For example, youth should be encouraged to critically evaluate their social networks by curating their online connections, limiting followers or friends to those who contribute positively to their digital environments, and establishing norms of respectful communication in group chats and forums. These actions foster psychological safety and create an environment for more genuine self-expression. EI activities coupled with digital literacy and human development education, especially identity formation, can help youth develop resilience.

4.3 Opportunities for positive relationships and social acceptance

Considering social relationships provide the social scaffolding necessary for adolescents to explore their values, beliefs, and roles, ultimately constructing a coherent sense of self, [25] it is essential that youth develop positive relationships in their lives.

4.3.1 Education/training for support networks

Educators, parents, and community members benefit from cultivating effective communication and mentorship skills that foster youths' identity formation. To

serve as a supportive mentor, they can learn strategies for non-judgmental listening, encouragement, and validation [80]. Rogers [62] believed that people need unconditional positive regard from others, and feeling fully accepted leads to greater self-acceptance and authenticity. Educational or therapeutic opportunities can be provided through online workshops, webinars, and community courses. For example, parent workshops can successfully help them learn to navigate this developmental time for youth [61].

4.3.2 Youth as positive feedback generators

Learning how to offer meaningful and supportive feedback can play a vital role in cultivating emotionally safe, inclusive, and flourishing online spaces. Waterman [32] underscored the developmental significance of feedback received from influential social agents, including parents, educators, coaches, and peers, in shaping identity and self-perception. While traditional focus has been placed on the influence of adults, peer feedback during adolescence is particularly salient, given its heightened relevance in the context of social comparison and identity formation. To that end, youth can be taught to engage more intentionally in their online interactions by moving beyond passive indicators of approval (e.g., likes) and instead offering specific, positive commentary that reinforces their peers' efforts, creativity, or values. For example, rather than simply liking a photo, a young person might comment on what they appreciated about the content, such as its message, artistic composition, or emotional tone. Or perhaps they can make a point to like and comment on 'authentic posts' supporting real and raw compositions from their peers, as more desirable. Such practices promote digital prosocial behavior, affirm authenticity, strengthen peer relationships, and foster a greater sense of belonging and digital well-being.

4.3.3 Limiting online social groups

Adolescents benefit from approaching their digital social networks with intentionality and critical awareness. Rather than indiscriminately expanding their online connections, youth should be encouraged to thoughtfully select digital friend groups based on mutual trust, respect, and shared values. Emerging research indicates that maintaining large online networks composed of individuals with whom one has weak or superficial ties increases the likelihood of encountering negative feedback, social comparison, and even online harassment or cyberbullying [54, 99]. By curating their digital environments to include close friends and supportive peers, youth can foster safer and more affirming online experiences that align with their identity development and psychological well-being. Teaching adolescents at home or in schools to critically assess who they allow into their digital spaces is, therefore, a crucial component of digital literacy and SEL skills.

4.3.4 Caregivers and educators as mentors and coaches

The adults in youths' lives occupy a critical position as mentors and coaches as adolescents explore identity development. Rather than viewing technology solely as a source of risk, adults should recognize its potential as a meaningful tool for identity exploration, social connection, and creative self-expression. When caregivers develop a nuanced understanding of the opportunities and challenges presented by digital

tools, such as gaming, social media, and online communities, they are better equipped to support youth in engaging with these platforms in developmentally constructive ways [100]. Caregivers can receive education about coaching and mentoring (e.g., parent webinars) and engage in collaborative technology use with their teens, which provides coaching opportunities.

A key aspect of coaching is learning how not to police youths' technology engagement. Youth need spaces to explore their emerging identities, especially those facing geographic or social isolation (e.g., living in rural communities or experiencing marginalization). Digital platforms often serve as a primary avenue for identity expression, peer feedback, and community building [101]. Consequently, punitive measures that involve removing access to technology (e.g., restricting smartphones or social media) can inadvertently disrupt critical developmental processes and lead to greater distress or social disconnection [102]. If there is a need to limit or remove digital tools (e.g., in response to online harm or during school-imposed phone bans), youth must be provided with alternative, developmentally appropriate avenues for self-expression and identity exploration. This might include opportunities in extracurricular clubs, arts-based programs, peer support spaces, or youth-led initiatives. In the context of increasing calls to regulate or restrict student access to personal devices in schools, Laffier [16] suggests that schools reimagine themselves as platforms for youth expression. Drawing inspiration from the features of social media, schools can incorporate strategies such as hallway murals serving as interactive backgrounds or comment walls for written peer-to-peer affirmations. These innovations preserve the social and expressive functions of digital platforms while promoting face-to-face connections, creativity, and a sense of belonging in the physical school environment.

4.4 Promotion of self-acceptance

Self-acceptance mediates authentic self-expression [57]. Therefore, youth need opportunities to express genuine aspects of themselves, along with the skills and encouragement to do so.

4.4.1 Creative pursuits

Self-expression is most developmentally meaningful when it involves a degree of authenticity and personal reflection, allowing individuals to engage with and articulate the positive aspects of themselves. In educational settings, assignments and activities should be intentionally designed to encourage students to reflect on their identities, strengths, and contributions. Such practices can facilitate self-awareness, reinforce a sense of personal agency, and support the development of a coherent and positive self-concept. The creative arts offer a psychologically safe and exploratory space for youth to engage in authentic self-expression [103]. Through modalities such as visual arts, music, drama, and creative writing, students can explore and represent dimensions of their identity in a manner that is both genuine and imaginative in a supportive and reflective context [103].

4.4.2 Practice gratitude and appreciation

Within psychological research, self-directed gratitude is increasingly recognized as a valuable self-care strategy that contributes to emotional well-being, self-compassion, and the development of a positive identity [104]. Gratitude, defined as the

recognition and appreciation of positive aspects of one's life and experiences, fosters reflective awareness and helps individuals focus on sources of personal meaning and strength [104]. Encouraging youth to identify and express appreciation for their personal qualities, experiences, and values may strengthen their self-concept and promote internal validation. Such practices shift attention from external standards or peer approval toward an appreciation of one's inherent worth and growth. This may encourage youth to focus on authentic aspects of themselves instead of hyper-idealized digital selves, which exacerbate vulnerabilities to mental health challenges [17]. Practical strategies for integrating gratitude into adolescents' daily routines include using gratitude journals, mobile apps designed for gratitude tracking, and reflective prompts facilitated by educators or parents.

4.5 Ethical and health-promotion technology design

Adolescents' relationship with technology should be understood as a form of social scaffolding that parallels the influence of parents and educators. Digital platforms have the capacity to shape learning, self-perception, and performance. Given this powerful influence, there is a pressing need for policy development that addresses digital feedback systems, usage limits, addictive design features, and monitoring practices.

4.5.1 AI and adolescent development

With the rapid advancement of artificial intelligence and its growing integration into adolescents' lives, it is essential to examine its impact on identity development and psychological well-being. While AI-enabled tools offer novel opportunities for self-exploration and identity experimentation, scholars and child advocacy organizations have emphasized the need for alignment with developmental theory, mental health frameworks, and ethical design principles. Without such considerations, AI-powered technologies risk undermining adolescent well-being during this critical stage of identity formation [16, 105].

4.5.2 Shifting the focus to health promotion

While existing discourse has largely focused on the risks and ethical concerns of youth engagement with technology, there is a need to expand this conversation to include the promotion of psychological health, including positive identity development. The U.S. Surgeon General's Advisory on Social Media and Youth Mental Health [106] recommends that technology companies design platforms that prioritize developmental benefits over maximizing time, attention, and engagement. A health promotion lens would consider how design choices can support emotional resilience, self-reflection, and authentic self-expression.

4.5.3 Designing for digital identity congruence

Technology design can play a vital role in fostering identity congruence by promoting authenticity and reflective self-expression. For example, BeReal is a social media platform that encourages users to share unscripted glimpses of their daily lives, aiming to counter the curated nature of traditional social media and foster genuine relationships [80]. Other platforms could adopt similar principles by integrating

features that promote authenticity, offer more nuanced emotional expressions, and encourage meaningful self-disclosure. Moving beyond binary reactions such as “like” or “dislike” could enable users to engage more deeply with themselves and others, aligning platform design with the developmental needs of youth.

5. Conclusion

This chapter examined the construct of digital identity congruence and its significance for adolescent mental health in the digital age. Adolescence is a period marked by identity exploration. As they increasingly turn to online spaces for self-expression and connection, the gap between their authentic offline selves and their curated digital personas has grown more pronounced. When these representations become detached from their authentic selves, especially without awareness or reflection, this incongruence can contribute to increased vulnerability to mental health challenges [17]. This concern is amplified by the pressures of algorithmic feedback, peer comparison, and the pursuit of digital acceptance [19].

Drawing on theoretical and empirical research, we argue that digital identity congruence, the alignment between one's digital and offline self, plays a critical role in supporting psychological well-being during this formative developmental stage [15, 16]. To address these challenges, we propose four key areas of intervention: (1) Positive Youth Development (PYD) approaches that empower identity formation, (2) supportive relationships that foster authenticity and social acceptance (3) promotion of self-awareness and self-acceptance as protective factors, and (4) technology design and policy focused on health promotion and congruent self-expression.

Ultimately, promoting reflective digital engagement, grounded in emotional intelligence and digital literacy, can enable youth to explore and express their identities online without compromising their mental health. Supporting digital identity congruence is not about limiting technological engagement but about reimagining digital spaces as platforms for authentic growth, resilience, and flourishing. These platforms can serve as both mirrors, reflecting adolescents' self-concepts, and as artistic canvases, allowing them to actively construct and express who they are or wish to become.

Author details

Jennifer Laffier* and Madison Westley
Ontario Tech University, Oshawa, Canada

*Address all correspondence to: jennifer.laffier@ontariotechu.ca

IntechOpen

© 2025 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Wilde O. *De Profundis*. In: Ellmann R, editor. *Original Work Written 1897*. London: Penguin Classics; 1959
- [2] Erikson EH. *Childhood and Society*. 2nd ed. New York: Norton; 1963
- [3] McFarlane K. Adolescent identity development in a digital world. *Journal of Adolescent Research*. 2024;**39**(1):34-52
- [4] van Doeselaar L, Becht AI, Klimstra TA, Meeus WHJ. A review and integration of three key components of identity development: Distinctiveness, coherence, and continuity. *European Psychologist*. 2018;**23**(4):278-288. DOI: 10.1027/1016-9040/a000334
- [5] Branje S, van Doeselaar L, Meeus W, Klimstra T. Identity development in adolescence: A decade in review. *Journal of Research on Adolescence*. 2021;**31**(4):908-927. DOI: 10.1111/jora.12678
- [6] Meeus W. The study of adolescent identity formation 2000-2010: A review of longitudinal research. *Journal of Research on Adolescence*. 2011;**21**(1):75-94. DOI: 10.1111/j.1532-7795.2010.00716.x
- [7] Branje S, Koper N. Identity development in adolescence and emerging adulthood. In: Padilla-Walker LM, Nelson LJ, editors. *Flourishing in Emerging Adulthood*. Oxford: Oxford University Press; 2018. pp. 125-145
- [8] Kroger J, Martinussen M, Marcia JE. Identity status change during adolescence and young adulthood: A meta-analysis. *Journal of Adolescence*. 2010;**33**(5):683-698. DOI: 10.1016/j.adolescence.2009.11.002
- [9] Nesi J. The impact of social media on youth mental health: Challenges and opportunities. *North Carolina Medical Journal*. 2020;**81**(2):116-121. DOI: 10.18043/ncm.81.2.116
- [10] Rodgers H, Lloyd-Evans EC. Intimate snapshots: TikTok, algorithm, and the recreation of identity. *Anthways*. 2021;**1**(1):1-17. DOI: 10.5281/zenodo.5515620
- [11] Telzer EH, Rogers CR, Do KT. Social influence on positive youth development: A developmental neuroscience perspective. *Advances in Child Development and Behavior*. 2017;**54**:215. DOI: 10.1016/bs.acdb.2017.10.003
- [12] Rogers C. A theory of therapy, personality, and interpersonal relationships: As developed in the client-centered framework. In: Koch S, editor. *Psychology: A Study of a Science, Formulations of the Person and the Social Context*. Vol. 3. New York: McGraw-Hill; 1959. pp. 184-256
- [13] Stephen S. Congruent functioning: The continuing resonance of Rogers' theory. *Person-Centered Experiential Psychother*. 2023;**22**(4):397-416. DOI: 10.1080/14779757.2022.2164334
- [14] Sheldon KM, Gunz A. Psychological needs as basic motives, not just experiential requirements. *Journal of Personality*. 2009;**77**(5):1467-1492
- [15] Laffier J. Online identities, congruence and the digital world. In: *Mental Health Week, Canadian Mental Health Association*. Toronto, ON; Jan 2016
- [16] Laffier J, Rehman A, Westley M. A conceptual framework to understand the relationships between digital

wellness and artificial intelligence. *Cyberpsychology, Behavior and Social Networking*. 2025;**28**(7):478-488. DOI: 10.1089/cyber.2024.0546

[17] Michikyan M, Subrahmanyam K, Dennis J. Can you tell who I am? Neuroticism, extraversion, and online self-presentation among young adults. *Computers in Human Behavior*. 2014;**33**:179-183. DOI: 10.1016/j.chb.2014.01.010

[18] Nesi J, Prinstein MJ. Using social media for social comparison and feedback-seeking: Gender and popularity moderate associations with depressive symptoms. *Journal of Abnormal Child Psychology*. 2015;**43**(8):1427-1438. Available from:. DOI: 10.1007/s10802-015-0020-0

[19] George MJ, Odgers CL. Seven fears and the science of how mobile technologies may be influencing adolescents in the digital age. *Perspectives on Psychological Science*. 2015;**10**(6):832-851. Available from:. DOI: 10.1177/1745691615596788

[20] World Health Organization. Adolescent Health [Internet]. Geneva, Switzerland: World Health Organization; 2024. Available from: <https://www.who.int/health-topics/adolescent-health/>

[21] UNESCO. Youth: Definitions and Engagement [Internet]. Paris, France: UNESCO; 2023. Available from: <https://www.unesco.org/en/youth>

[22] Baird S, Choonara S, Azzopardi PS, et al. A call to action: The second lancet commission on adolescent health and wellbeing. *Lancet Commissions*. 2025;**405**:1945-2022. DOI: 10.1016/S0140-6736(25)00503-3. Available from: <https://www.healthdata.org/research-analysis/library/call-action-second-lancet-commission-adolescent-health-and-wellbeing>

[23] Rehman A. An exploration of artistic creation in makerspaces and youth identity [thesis]. Oshawa (ON): University of Ontario Institute of Technology; 2020

[24] Sandberg DJ, Frisé A, Eriksson PL, Syed M. Personality traits as predictors of identity intervention outcomes in adolescence. *Journal of Youth and Adolescence*. 2025;**54**(3):412-430

[25] Eichas K, Côté JE, Levine CG. Identity Formation, Youth, and Development: A Simplified Approach. 2nd ed. New York: Routledge; 2025

[26] Manago AM. Media and the development of identity. In: Scott RA, Kosslyn SM, editors. *Emerging Trends in the Social and Behavioral Sciences*. Hoboken (NJ): Wiley; 2015. pp. 1-15. DOI: 10.1002/9781118900772.etrds0212

[27] Chen K, Lay K, Wu Y, Yao G. Adolescent self-identity and mental health: The function of identity importance, identity firmness, and identity discrepancy. *Chinese Journal of Psychology*. 2007;**49**(1):53

[28] Sus V. Psychosocial Development Theory Explained [Internet]. Vancouver, British Columbia, Canada: Helpful Professor; 2024. Available from: <https://helpfulprofessor.com/psychosocial-development-theory-erikson-stages/>

[29] Marcia JE. Development and validation of ego-identity status. *Journal of Personality and Social Psychology*. 1966;**3**(5):551-558

[30] Arnold ME. Supporting adolescent exploration and commitment: Identity formation, thriving, and positive youth development. *Journal of Youth Development*. 2017;**12**(4):5

[31] Crocetti E. Identity formation in adolescence: The dynamic of forming

- and consolidating identity commitments. *Child Development Perspectives*. 2017;**11**(2):145-150. DOI: 10.1111/cdep.12226
- [32] Waterman AS. Personal expressiveness: Philosophical and psychological foundations. *Journal of Mind & Behavior*. 1990;**11**(1):47-73
- [33] Crocetti E, Meeus W. Identity formation and adjustment in adolescence. In: Monteiro MJ, Soares M, editors. *Handbook of Research on Individualism and Identity in the Globalized Digital Age*. Hershey, Pennsylvania, USA: IGI Global; 2015. pp. 1-16
- [34] Stets JE, Burke PJ. Identity nonverification, emotions, and cognitive and behavioral responses. *Social Psychology Quarterly*. 2025:1-25. DOI: 10.1177/01902725251318211
- [35] Gonzales AL, Hancock JT. Identity shift in computer-mediated environments. *Media Psychology*. 2008;**11**(2):167-185. DOI: 10.1080/15213260802023433
- [36] Anderson M, Faverio M, Park E. How Teens and Parents Approach Screen Time. Washington, DC, USA: Pew Research Center; 2024. Available from: <https://www.pewresearch.org/internet/2024/03/11/how-teens-and-parents-approach-screen-time/#teens-views-on-screen-time-and-efforts-to-cut-back>
- [37] Faverio M, Sidoti O. Teens, Social Media and Technology, 2024. Washington, DC, USA: Pew Research Center; 2024. Available from: <https://www.pewresearch.org/internet/2024/12/12/teens-social-media-and-technology-2024/>
- [38] Pérez-Torres V. Social media: A digital social mirror for identity development during adolescence. *Current Psychology*. 2024;**43**:22170-22180. DOI: 10.1007/s12144-024-05980-z
- [39] Raiziene S, Erentaite R, Pakalniskiene V, Grigutyte N, Crocetti E. Identity formation patterns and online activities in adolescence. *Identity*. 2021;**22**(2):150-165. DOI: 10.1080/15283488.2021.1960839
- [40] Straight S, Abrams Z. *How Psychology Is Shaping the Future of Technology*. Washington, DC, USA: American Psychological Association; 2024. Available from: <https://www.apa.org/news/apa/2024/psychology-shape-future-technology> [Accessed: July 9, 2025]
- [41] Waterman AS. Identity formation: Discovery or creation? *Journal of Early Adolescence*. 1984;**4**(4):329-341. DOI: 10.1177/0272431684044004
- [42] Gottfried J, Sidoti O. *Teens and Video Games Today*. Washington, DC, USA: Pew Research Center; 2024. Available from: <https://www.pewresearch.org/internet/2024/05/09/teens-and-video-games-today/>
- [43] Szolin K, Kuss DJ, Nuyens FM, Griffiths MD. “I am the character, the character is me”: A thematic analysis of the user-avatar relationship in videogames. *Computers in Human Behavior*. 2023;**143**:107694
- [44] Tichon J, Makaresz S. Influencing young adult social and personal identity through video-game narratives. *Asia Pacific Journal of Advanced Business and Social Studies*. 2019;**5**(1):116-129
- [45] Wood SM, Szymanski A. “The me I want you to see”: The use of video game avatars to explore identity in gifted adolescents. *Gifted Child Today*. 2020;**43**(2):124-134

- [46] Labadze L, Grigolia M, Machaidze L. Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*. 2023;**20**(1):56
- [47] Valkenburg PM, Piotrowski JT. *Plugged in: How Media Attract and Affect Youth*. New Haven (CT): Yale University Press; 2017
- [48] Laffier J, Westley M. Young adults, movies, and trauma: Educating pre-service teachers about trauma-informed practices. In: *Proceedings of the SITE Conference*, San Diego, CA. Waynesville, North Carolina, USA: Association for the Advancement of Computing in Education (AACE); 2022
- [49] DeAngelis T. Teens Are Spending Nearly 5 Hours Daily on Social Media: Here Are the Mental Health Outcomes. Washington, DC, USA: APA Monitor; 2024. Available from: <https://www.apa.org/monitor/2024/04/teen-social-use-mental-health>
- [50] González-Larrea B, Hernández-Serrano MJ, Morales RN. Online identity construction in younger generations via identification with influencers: Potential areas of vulnerability. In: *Identity in a Hyperconnected Society*. Cham: Springer; 2021. pp. 95-107. DOI: 10.1007/978-3-030-85788-2_7
- [51] Bates A, Hobman T, Bell BT. "Let me do what I please with it... don't decide my identity for me": LGBTQ+ youth experiences of social media in narrative identity development. *Journal of Adolescent Research*. 2020;**35**(1):51-83. DOI: 10.1177/0743558419884700
- [52] Craig SL, McInroy L, McCready LT, Alaggia R. Media: A catalyst for resilience in lesbian, gay, bisexual, transgender, and queer youth. *Journal of LGBT Youth*. 2015;**12**(3):254-275
- [53] Broz M. *Selfie Statistics, Demographics, and Fun Facts* [Internet]. Slovenia: Photutorial; 2025. Available from: <https://photutorial.com/selfie-statistics/>
- [54] Kowalski RM, Giumetti GW, Schroeder AN, Lattanner MR. Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*. 2014;**140**(4):1073-1137. DOI: 10.1037/a0035618
- [55] Mann RB, Blumberg F. Adolescents and social media: The effects of frequency of use, self-presentation, social comparison, and self-esteem on possible self imagery. *Acta Psychologica*. 2022;**228**:103629. DOI: 10.1016/j.actpsy.2022.103629
- [56] SAMHSA Kids Online Health and Safety Task Force. *Online Health and Safety for Children and Youth: Best Practices for Families and Guidance for Industry* [Internet]. Gaithersburg (MD): Substance Abuse and Mental Health Services Administration; 2024. Report No.: PEP24-01-012. Available from: <https://www.samhsa.gov/sites/default/files/online-health-safety-children-youth-report.pdf>
- [57] Bailey ER, Matz SC, Youyou W, Iyengar SS. Authentic self-expression on social media is associated with greater subjective well-being. *Nature Communications*. 2020;**11**(1):4889. DOI: 10.1038/s41467-020-18639-w
- [58] Manago AM, Perez de la Torre MM, de la Torre Sánchez MC. Cultural identity as worldviews: A natural experiment with Maya adolescents before and after community adoption of digital communication. *Developmental Psychology*. 2025:1-18. DOI: 10.1037/dev0001674

- [59] Erikson EH. *Identity: Youth and Crisis*. New York: W. W. Norton & Company; 1968
- [60] Schwartz R, Vassilev A, Greene K, Perine L, Burt A, Hall P. Towards a Standard for Identifying and Managing Bias in Artificial Intelligence [Internet]. Gaithersburg (MD): National Institute of Standards and Technology (U.S.); 2022. Report No.: NIST SP 1270. Available from: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf>
- [61] Soh S, Cruz B, Meca A, Harari GM. A digital identity intervention incorporating social media activities promotes identity exploration and commitment among emerging adults. *Identity*. 2024;**25**(2):330-346. DOI: 10.1080/15283488.2024.2404004
- [62] Rogers C. *On Becoming a Person: A therapist's View of Psychotherapy*. Boston: Houghton Mifflin; 1961
- [63] Rogers CR. The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*. 1957;**21**(2):95-103. DOI: 10.1037/h0045357
- [64] Bleuler E. *Die Ambivalenz* [Internet]. Zürich, Switzerland: Universität Zürich; 1910. Available from: <http://www.sgipt.org/medppp/gesch/ambiv-g.htm>
- [65] Buttler B, Pauer S, van Harreveld F. The model of ambivalent choice and dissonant commitment: An integration of dissonance and ambivalence frameworks. *European Review of Social Psychology*. 2024;**36**(1):195-237. DOI: 10.1080/10463283.2024.2373547
- [66] Laffier J. Depression and suicide in the virtual worlds of youth. In: Lester D, editor. *Understanding Suicide Ideation*. New York: Nova Science Publishers; 2020. pp. 137-165. Available from: <https://psycnet.apa.org/record/2020-38551-008>
- [67] Rudd MD. *Assessment and Management of Suicidality*. Sarasota (FL): Professional Resource Press; 2006
- [68] Festinger L. Cognitive dissonance. *Scientific American*. 1962;**207**(4):93-106
- [69] Al-Adamat OA, Atoum AY. Cognitive dissonance and its relationship to emotional intelligence. *Cognition Brain Behavior*. 2022;**26**(4):215-229
- [70] Dilakshini VL, Kumar SM. Cognitive dissonance: a psychological unrest. *Current Journal of Applied Science and Technology*. 2020;**39**(30):54-60
- [71] Cornelius-White JH. Congruence: an integrative five-dimension model. *Person-Centered and Experiential Psychotherapies*. 2007;**6**(4):229-239
- [72] Bozarth JD, Wilkins P, editors. *Rogers' Therapeutic Conditions: Evolution, Theory and Practice*. Vol. 3, Unconditional Positive Regard. Ross-on-Wye (UK): PCCS Books; 2001
- [73] Gardner H. *The App Generation: How today's Youth Navigate Identity, Intimacy, and Imagination in a Digital World*. New Haven (CT): Yale University Press; 2013
- [74] Altman I, Taylor DA. *Social Penetration: The Development of Interpersonal Relationships*. New York: Holt, Rinehart & Winston; 1973
- [75] Sala A, Porcaro L, Gómez E. Social media use and adolescents' mental health and well-being: An umbrella review. *Computers in Human Behavior Reports*. 2024;**14**:100404
- [76] Pascual-Leone A. How clients "change emotion with emotion"

: A programme of research on emotional processing. *Psychotherapy Research*. 2018;**28**(2):165-182.
 DOI: 10.1080/10503307.2017.1349350

[77] Avci H, Baams L, Kretschmer T. A systematic review of social media use and adolescent identity development. *Adolescent Research Review*. 2025;**10**(2):219-236. DOI: 10.1007/s40894-024-00251-1

[78] Schlenker BR. Self-presentation. In: Leary MR, Tangney JP, editors. *Handbook of Self and Identity*. 2nd ed. New York: Guilford Press; 2003. pp. 542-570

[79] Deeb-Swihart J, Polack C, Gilbert E, Essa I. Selfie-presentation in everyday life: A large-scale characterization of selfie contexts on Instagram. In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 11(1). Montréal, Québec, Canada: Association for the Advancement of Artificial Intelligence; 2017. pp. 42-51

[80] Kim J, Wolfe R, Chordia I, Davis K, Hiniker A. "Sharing, not showing off": How BeReal approaches authentic self-presentation on social media through its design. In: *Proc ACM Hum-Comput Interact*. Vol. 8. 2024. pp. 1-32. DOI: 10.1145/3686909

[81] Ryan RM. *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. New York: Guilford Publications; 2017

[82] Mace R. A real virtual self. In: *Children, Young People and Online Harms*. Cham: Palgrave Macmillan; 2024. pp. 35-60. DOI: 10.1007/978-3-031-46053-1_2

[83] Lemenager T, Neissner M, Sabo T, Mann K, Kiefer F. "Who am I" and "how should I be": A systematic review on

self-concept and avatar identification in gaming disorder. *Current Addiction Reports*. 2020;**7**(2):166-193

[84] Xu K, Geng S, Dou D, Liu X. Relations between video game engagement and social development in children: The mediating role of executive function and age-related moderation. *Behavioral Sciences (Basel, Switzerland)*. 2023;**13**(10):833. DOI: 10.3390/bs13100833

[85] Forsyth SR, Chesla CA, Rehm RS, Malone RE. "It feels more real": An interpretive phenomenological study of the meaning of video games in adolescent lives. *Advances in Nursing Science*. 2017;**40**(4):E1-E17. DOI: 10.1097/ANS.0000000000000183

[86] Anli G. Examining the predictive role of authenticity on internet addiction in Turkish high school students. *Universal Journal of Educational Research*. 2018;**6**(7):1497-1503. DOI: 10.13189/ujer.2018.060710

[87] Goldner L, Jakobi CD, Schorr S, Dakak S, Shawahne N. Keep it quiet: Mother-daughter parentification and difficulties in separation-individuation shaping daughters' authentic/true self and self-silencing: A mediation model. *Psychoanalytic Psychology*. 2022;**39**(2):165-175

[88] Alchin CE, Machin TM, Martin N, Burton LJ. Authenticity and inauthenticity in adolescents: A scoping review. *Adolescent Research Review*. 2024;**9**(2):279-315

[89] Luthar SS, Ebbert AM, Kumar NL. Risk and resilience among Asian American youth: Ramifications of discrimination and low authenticity in self-presentations. *The American Psychologist*. 2021;**76**(4):643-657. DOI: 10.1037/amp0000764

- [90] Nartova-Bochaver SK, Bayramyan RM, Chulyukin KS, Yerofeyeva VG. What it means to be oneself: The everyday ideas of authenticity among primary school children and adolescents in Russia. *Psychology in Russia : State of the Art.* 2021;**14**(3):3-21. DOI: 10.11621/pir.2021.0301
- [91] Lerner RM, Lerner JV, Geldhof GJ, Gestsdóttir S, King PE, Sim AT, et al. Studying positive youth development in different nations. In: Schwartz SJ, Luyckx K, Vignoles VL, editors. *Handbook of Adolescent Development Research and its Impact on Global Policy.* Cham (Switzerland): Springer; 2018. pp. 68-83
- [92] Liverpool S, McDonagh C, Feather J, Mateus C. Updates on digital mental health interventions for children and young people: A systematic overview. *European Child & Adolescent Psychiatry.* 2025;**34**(2):145-162. DOI: 10.1007/s00787-025-02722-9
- [93] Eichas K, Montgomery MJ, Meca A, Kurtines WM. Empowering marginalized youth: A self-transformative intervention for promoting positive youth development. *Child Development.* 2017;**88**(4):1115-1124. DOI: 10.1111/cdev.12866
- [94] Umaña-Taylor AJ, Douglass S, Updegraff KA, Marsiglia FF. A small-scale randomized efficacy trial of the identity project: Promoting adolescents' ethnic-racial identity exploration and resolution. *Child Development.* 2018;**89**(3):862-870. DOI: 10.1111/cdev.12753
- [95] Perkins DD, Zimmerman MA. Empowerment theory, research, and application. *American Journal of Community Psychology.* 1995;**23**(5):569-579. DOI: 10.1007/BF02506982
- [96] Neff KD. Self-compassion, self-esteem, and well-being. *Social and Personality Psychology Compass.* 2011;**5**(1):1-12. DOI: 10.1111/j.1751-9004.2010.00330.x
- [97] Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist.* 2000;**55**(1):68-78. DOI: 10.1037/0003-066X.55.1.68
- [98] Collaborative for Academic, Social, and Emotional Learning (CASEL). CASEL's SEL Framework: What Are the Core Competence Areas and Where Are they Promoted? Chicago, IL: CASEL; 2020. Available from: <https://casel.org/casel-sel-framework-11-2020/>
- [99] Marwick AE, boyd D. Networked privacy: How teenagers negotiate context in social media. *New Media & Society.* 2014;**16**(7):1051-1067. DOI: 10.1177/1461444814543995
- [100] Odgers CL, Jensen MR. Annual research review: Adolescent mental health in the digital age: Facts, fears, and future directions. *Journal of Child Psychology and Psychiatry.* 2020;**61**(3):336-348. DOI: 10.1111/jcpp.13190
- [101] Best P, Manktelow R, Taylor B. Online communication, social media and adolescent wellbeing: A systematic narrative review. *Children and Youth Services Review.* 2014;**41**:27-36. DOI: 10.1016/j.childyouth.2014.03.001
- [102] Uhls YT, Michikyan M, Morris J, Garcia D, Small GW, Zgourou E, et al. Five days at outdoor education camp without screens improves preteen skills with nonverbal emotion cues. *Computers in Human Behavior.* 2014;**39**:387-392. DOI: 10.1016/j.chb.2014.05.036

[103] Rubin JA. Approaches to Art Therapy: Theory and Technique. 2nd ed. New York: Routledge; 2012

[104] Emmons RA, Stern R. Gratitude as a psychotherapeutic intervention. *Journal of Clinical Psychology*. 2013;**69**(8):846-855. DOI: 10.1002/jclp.22020

[105] eSafety Commissioner. AI Chatbots and Companions – Risks to Children and Young People [Internet]. Australia: eSafety Commissioner, Australian Government; 2025. Available from: <https://www.esafety.gov.au/newsroom/blogs/ai-chatbots-and-companions-risks-to-children-and-young-people>

[106] U.S. Department of Health and Human Services. Office of the Surgeon General. Surgeon General's Advisory on Social Media and Youth Mental Health. Washington (DC): U.S. HHS; 2023. Available from: <https://www.hhs.gov/about/news/2023/05/23/surgeon-general-issues-new-advisory-about-effects-social-media-use-has-youth-mental-health.html>

Chapter 4

EdgeAI with TinyML: Redefining Privacy and Security in Online Identity Management

Yogeswar Reddy Thota and Tooraj Nikoubin

Abstract

As online identity management evolves, traditional cloud-based authentication methods are facing critical challenges related to privacy, latency, and centralized risk. This chapter explores how Tiny Machine Learning (TinyML) and Edge Artificial Intelligence (EdgeAI) enable real-time, on-device biometric authentication to address these challenges. TinyML models deployed on ultra-low-power devices such as smartphones, wearables, and IoT (Internet of Things) sensors allow biometric data such as facial recognition, voice, and physiological signals like ECG (electrocardiogram) and PPG (photoplethysmogram) to be processed locally, reducing dependency on cloud infrastructure and minimizing data exposure. We explore the use of ephemeral cryptographic tokens to support passwordless authentication and the role of secure model deployment, Zero Trust Architecture (ZTA), and federated learning in protecting identity across edge ecosystems. The chapter also highlights real-world use cases from healthcare, mobile access, and border security, and addresses ethical and regulatory challenges including data consent, demographic bias, and compliance with laws like the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA). By offering a comprehensive review of decentralized identity frameworks, This chapter contributes a comprehensive systematization of how TinyML and EdgeAI are applied to real-time, privacy-preserving identity verification, along with security architecture, use cases, and regulatory framing.

Keywords: TinyML, EdgeAI, online identity, passwordless authentication, privacy-preserving AI

1. Introduction

In today's increasingly digital society, the concept of online identity has evolved from simple username-password mechanisms to complex systems involving biometric verification, behavioral analysis, and multi-device interactions. As our dependence on digital infrastructure grows, so too do the risks associated with centralized identity storage and legacy authentication protocols. Data breaches, phishing attacks, credential stuffing, and biometric spoofing have become common, with cybercriminals targeting the digital identities of individuals and organizations

alike [1, 2]. These threats are further exacerbated by the explosive growth of the Internet of Things (IoT), wearables, and edge-connected systems, many of which operate under severe resource constraints. Such devices often lack the computational power and memory to support traditional security frameworks, making them attractive targets for adversaries [3]. Additionally, reliance on cloud-based processing introduces latency, data sovereignty concerns, and privacy vulnerabilities due to the transmission and storage of sensitive biometric or behavioral data on remote servers [4].

Online identity management systems have traditionally relied on centralized cloud infrastructure, which, while scalable, presents a single point of failure in terms of security and user privacy. Prior studies such as [5] have proposed cryptographic and [6] multi-factor methods for improving identity security, but often lack real-time adaptability and on-device autonomy. Moreover, recent frameworks for privacy-preserving identity, such as self-sovereign identity (SSI) models [7, 8], remain largely theoretical or dependent on complex blockchain integration posing adoption barriers in low-power or real-time contexts. Study at [9, 10] demonstrate EdgeAI-based biometric authentication and template protection using trusted enclaves and lightweight models. However, a unified systematization linking nano-scale identity inference with regulatory, architectural, and ethical dimensions remains absent.

To address these challenges, emerging technologies such as Tiny Machine Learning (TinyML) and Edge Artificial Intelligence (EdgeAI) offer a compelling alternative. TinyML refers to the deployment of optimized machine learning models on ultra-low-power devices, such as microcontrollers and embedded processors. These models are capable of performing inference locally, eliminating the need to transmit raw data to external servers. EdgeAI complements this approach by enabling intelligent, context-aware decision-making directly on distributed devices, enhancing system responsiveness, privacy, and robustness [11, 12].

This shift from cloud-centric identity verification to on-device, decentralized authentication represents a fundamental rethinking of online identity. In this model, user verification can occur through local biometric processing such as facial recognition, voice matching, or behavioral signatures followed by the generation of an ephemeral, cryptographically secure token that is used to authenticate with remote services. This not only improves speed and efficiency but also reduces the exposure of sensitive credentials to the network [13].

Furthermore, technologies like federated learning enable continuous improvement of identity models across multiple user devices without sharing raw data. This approach aligns with global privacy regulations such as the General Data Protection Regulation (GDPR), while maintaining the adaptability of AI systems to new threats and usage patterns [14]. Additional safeguards such as Trusted Execution Environments (TEEs) and adversarially robust models provide further layers of protection for on-device inference [15, 16].

Despite these advances, a consolidated and systematized perspective on how TinyML and EdgeAI technologies coalesce to support privacy-preserving online identity is missing from current literature. This chapter aims to fill that gap by presenting a unified framework that addresses technical, regulatory, and ethical dimensions of decentralized identity using real-time, edge-native architectures. In the following sections, we will examine their role in biometric-based identity inference, passwordless login systems, federated identity learning, and edge-native threat detection. Real-world use cases, system architectures, and regulatory considerations will be discussed to provide a comprehensive roadmap for secure digital identity at the edge. The next

sections delve into how TinyML specifically supports decentralized identity verification through various biometric and token-based methods.

2. TinyML for biometric identity inference

Building upon the foundation of decentralized identity, this section explores how TinyML enables localized biometric verification on edge devices. The rising adoption of biometric systems such as fingerprint scanners, facial recognition, voice authentication, and behavioral biometrics has marked a significant shift from knowledge-based authentication (e.g., passwords) to identity-based access control. However, conventional biometric verification systems typically rely on powerful cloud infrastructure for feature extraction, model inference, and authentication decision-making. This introduces latency, poses privacy concerns due to the transmission of sensitive data, and increases vulnerability to interception or spoofing during transit [4].

Tiny Machine Learning (TinyML) offers a compelling solution by enabling lightweight machine learning models to run entirely on edge devices, including microcontrollers, IoT nodes, wearables, and smartcards [11]. By processing biometric data locally, TinyML systems eliminate the need to transmit raw features to centralized servers. This architectural shift reduces both bandwidth requirements and potential exposure of personally identifiable information (PII), while maintaining real-time responsiveness in constrained environments [11, 12]. An illustration of Biometric signal based local verification flow using TinyML model is shown in **Figure 1**.

A key advantage of TinyML in the context of online identity is its ability to support on-device biometric inference. For example, a smart wristband could continuously monitor the user's heart rhythm or motion patterns using accelerometer and photoplethysmography (PPG) sensors. A trained TinyML model, deployed directly on the device, can verify the biometric signal against a stored template and confirm user presence or identity without relying on an external cloud service [3, 11].

Several modalities of biometric verification can benefit from TinyML-based local inference, comparison of these Biometric Modalities are listed in **Table 1**.

- Facial recognition: Tiny convolutional neural networks (CNNs) like MobileNetV2 or SqueezeNet can be compressed and quantized to run on ARM Cortex-M devices or similar microcontrollers [11, 17].
- Voice authentication: Keyword spotting or speaker verification using models like DS-CNN and GRU architectures can be optimized for low-power audio chips, enabling voice-based authentication on smart speakers or earbuds [18].

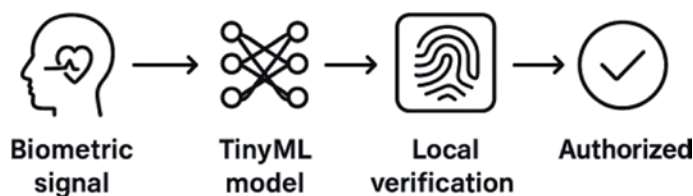


Figure 1.
Biometric signal based local verification flow using TinyML model.

Biometric modality	Signal type	Device type	Privacy level	Attack resistance	Power requirement
Face Recognition	Visual (2D/3D)	Smartphone, Laptop	Medium	Medium	Moderate
Voice Authentication	Audio Waveform	Earbuds, Speakers	Medium	Low (Replay Risk)	Low
Keystroke Dynamics	Typing Behavior	Keyboard, Laptop	High	High	Very Low
Gait Recognition	Motion/ IMU	Wearables	High	Medium	Low
ECG (Heart Rhythm)	Bioelectric Signal	Smartbands	Very High	Very High	Moderate
PPG (Pulse Waveform)	Optical Signal	Fitness Bands	Very High	High	Low

Table 1.
Comparison of biometric modalities supported by TinyML.

- Behavioral biometrics: TinyML models can analyze keystroke dynamics, gait, or touchscreen swiping patterns to establish behavioral signatures unique to each user [13, 19].
- Physiological biometrics (ECG/PPG): Signals like electrocardiograms (ECG) and PPG offer distinct physiological patterns that are difficult to replicate, making them highly secure for identity verification.

TinyML enables on-device processing of these waveforms using lightweight signal processing and classification models. Wearables equipped with heart rate sensors or bio-potential electrodes can perform user identification continuously and passively, without transmitting sensitive waveform data to external servers [20, 21]. Despite these advantages, deploying biometric inference at the edge is not without challenges. Devices are susceptible to spoofing attacks (e.g., voice replay, 3D face masks) and adversarial inputs designed to fool models into misclassification. To counter these threats, TinyML models must be trained with diverse datasets and augmented using adversarial defense techniques such as adversarial training, autoencoder filtering, or signal anomaly detection [16].

Furthermore, the integrity of model inference on edge devices must be protected through TEEs or secure boot mechanisms. These ensure that models cannot be tampered with or extracted by malicious actors [15]. In practical deployments, manufacturers often combine secure enclaves with on-device biometric modules, as seen in systems like Apple FaceID and Android’s Trusted Face implementations, though these often rely on proprietary architectures [12].

As resource-efficient architectures like TensorFlow Lite Micro, CMSIS-NN, and TinyMLPerf continue to evolve, more complex biometric models can be deployed securely and reliably on constrained devices. This advancement positions TinyML as a foundational pillar for privacy-preserving, scalable identity inference in a decentralized ecosystem. Extending the concept of on-device biometric verification, we now examine how ephemeral tokens can enable secure, passwordless authentication in next section.

3. Passwordless login using ephemeral tokens

The reliance on traditional usernames and passwords has long been a weak link in digital security. Password-based systems are vulnerable to phishing, credential stuffing, brute-force attacks, and poor user password hygiene [2]. Even when multi-factor authentication (MFA) is deployed, it typically depends on centralized architectures or user-dependent steps, both of which introduce friction and privacy concerns. Illustration of TinyML-powered token authentication flow is shown in **Figure 2**. A more secure and frictionless alternative lies in passwordless authentication, powered by TinyML-driven biometric verification and ephemeral cryptographic tokens. In this paradigm, identity is not confirmed through memorized knowledge, but through real-time validation of the user's biometric or behavioral traits, processed directly on the edge device. Once the biometric signal is verified locally by a TinyML model be it a face, voiceprint, keystroke pattern, or physiological signal a secure one-time token is generated. This ephemeral token acts as a cryptographic proof of identity, bound to both the device and the session. It is transmitted to the service provider (e.g., a login API) for verification, eliminating the need to transmit raw biometric data or reuse static credentials [13].

Recent studies have expanded the scope of biometric signals beyond conventional modalities, advancing toward physiological signals such as ECG, PPG, and other vital signs. These biosignals, captured from chest or wrist-worn sensors, offer highly individual patterns that are difficult to forge, making them ideal candidates for high-security applications [3]. With TinyML, such signals can be preprocessed and analyzed directly on microcontroller-class devices, enabling passive and continuous identity assurance.

This approach offers several critical advantages:

- Privacy-preserving: No raw biometric data is ever transmitted.
- Replay-resistant: Tokens are session-bound and cryptographically signed.
- Device-tied: Tokens are often generated using hardware root-of-trust elements or secure enclaves [15].
- Frictionless UX: Users do not need to remember or type credentials.

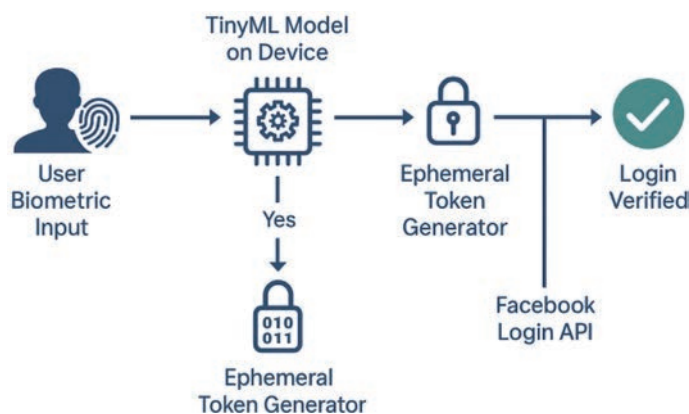


Figure 2.
TinyML-powered token authentication flow.

Technically, this process is supported by emerging protocols such as WebAuthn and FIDO2, which enable browser and platform-level support for biometric authentication without passwords. Companies like Apple, Google, and Microsoft have already begun rolling out passkeys a consumer-facing version of this paradigm relying on a combination of on-device biometrics and public-key cryptography.

An example flow is as follows:

- User initiates login on a web service.
- Device captures a biometric signal (e.g., face, PPG waveform).
- TinyML model performs real-time local verification.
- If successful, an ephemeral token is generated and signed.
- The token is transmitted to the authentication server.
- Server verifies token origin and integrity → access granted.

The tight integration between TinyML inference and cryptographic token generation allows even resource-constrained edge devices to become secure identity agents in a decentralized architecture. These systems can support a variety of deployment scenarios, including:

- Offline authentication at kiosks or field stations with intermittent internet.
- Multi-device identity fusion, where a smartwatch and smartphone jointly confirm identity via federated decision-making [14].
- Secure IoT access in industrial or medical settings without requiring human intervention.

As more services shift toward zero-trust architectures, this model of biometrically verified, token-based passwordless login offers a scalable, user-friendly, and secure alternative that aligns with both technical and regulatory demands. To ensure the reliability and security of TinyML-based identity systems, it is essential to protect the entire model lifecycle—from deployment to inference, which we illustrated in Section 4.

4. Securing the TinyML model lifecycle

As TinyML-based identity systems become more widely adopted, the need to secure the entire model lifecycle from training to deployment and inference becomes critical. Unlike traditional ML deployments that benefit from server-grade protections, TinyML models operate on resource-constrained, easily accessible devices, increasing the risk of tampering, reverse engineering, and adversarial exploitation.

4.1 Secure model deployment

Model integrity begins with protected deployment. When TinyML models are uploaded to a device, they must be transferred and stored in a way that prevents unauthorized access or modification. This is often achieved using:

- Encrypted firmware packaging, ensuring confidentiality of model weights.
- Digital signatures for verifying authenticity and integrity before execution.
- Secure bootloaders, which validate the model and device firmware before allowing inference.

Many modern microcontrollers and System-on-Chip (SoC) platforms support TEEs or Secure Boot chains that isolate model inference to protected memory regions inaccessible from the general application space [15].

4.2 Over-the-air (OTA) updates

A key advantage of TinyML in identity systems is the ability to adapt to user behavior or update threat models over time. However, this requires secure OTA model updates, which must be cryptographically signed, authenticated, and verified before installation. Weaknesses in the update mechanism can lead to model injection attacks, where adversaries replace models with compromised versions that either bypass security checks or leak sensitive information [16].

The update protocol must therefore:

- Use end-to-end encryption and mutual authentication.
- Validate model origin, integrity, and versioning.
- Enforce rollback protection and tamper evidence.

4.3 Protection against model extraction

Because TinyML models are deployed directly to physical devices, they are susceptible to model extraction attacks, where attackers gain access to the model and reconstruct it through:

- Firmware dumping,
- Side-channel analysis,
- Input-output queries combined with membership inference or model inversion techniques [16, 22].

To mitigate these threats:

- Sensitive biometric templates should not be embedded directly in model parameters.

- Lightweight obfuscation and quantization can increase extraction complexity.
- Hardware security modules (HSMs) or trusted enclaves can be used to isolate and encrypt model data.

In addition, researchers have proposed white-box model watermarking, where developers embed hidden patterns that prove model ownership and allow tamper detection if the model is stolen or reused [22].

4.4 On-device inference integrity and robustness

Once deployed, the TinyML model must maintain inference integrity against malicious inputs. This includes:

- *Adversarial input detection*, using threshold-based filters or lightweight anomaly detection autoencoders [16].
- *Redundant signal verification*, where multiple biometric modalities or time-series consistency checks are used to confirm identity.
- *Fail-closed behavior*, where authentication is denied by default when signals are inconclusive or model trust is compromised.

As biometric systems expand into passive and continuous verification modes (e.g., ongoing heartbeat or movement analysis), ensuring inference accuracy and trustworthiness becomes a cornerstone of identity security. By securing each stage of the TinyML lifecycle, from training and deployment to updates and runtime protection, developers can ensure that decentralized identity systems remain resilient, tamper-resistant, and privacy-preserving.

5. Integration with zero trust architecture

Complementing model security, the Zero Trust paradigm demands continuous verification, which TinyML and EdgeAI can fulfill through real-time risk analysis. In an era where identity-based attacks dominate cybersecurity breaches, organizations are rapidly adopting the Zero Trust Architecture (ZTA) framework. At its core, ZTA operates on a simple yet powerful principle: “never trust, always verify.” Rather than assuming implicit trust based on device location, credentials, or network perimeter, ZTA mandates continuous verification of identity, device posture, and behavioral context before granting or maintaining access [23]. TinyML and EdgeAI provide the technical foundation to operationalize this principle in real time and at scale. By enabling on-device identity verification and risk scoring, TinyML models can continuously evaluate whether a user is still authorized not just at login, but throughout an active session.

5.1 Continuous identity verification at the edge

Traditional identity systems authenticate users once per session, relying on static factors such as passwords or one-time codes. In contrast, TinyML allows for continuous, context-aware authentication by:

- Monitoring behavioral biometrics (e.g., typing rhythm, touchscreen gestures),
- Analyzing physiological patterns (e.g., heart rate variability via PPG/`111,111),
- Tracking device interactions (e.g., motion sensors, usage patterns).

For instance, a TinyML model deployed on a smart ring could passively verify a user's gait and heart rhythm in real-time. Any deviation from the known biometric profile can trigger a local risk score, which is then used to either maintain or revoke access privileges [3]. This granular approach to identity assurance minimizes the risk of session hijacking, credential sharing, or device theft-based access, aligning directly with ZTA principles of ongoing trust evaluation. **Figure 3** shows the flow of continuous risk scoring or micro-authentication loop.

5.2 Local risk scoring before cloud access

One of the core components of ZTA is the use of risk-based access control. TinyML can pre-process contextual risk signals such as abnormal usage time, geographic inconsistency, or biometric mismatch before allowing requests to be sent to the cloud or internal servers [24]. Instead of relying solely on cloud-based decision engines, this approach empowers the edge device to become a first-line policy enforcer, reducing latency and conserving bandwidth. Only if the risk score is within a trusted threshold is access permitted; otherwise, the request can be escalated for multi-factor verification or quarantined.

5.3 Micro-authentication events across sessions

TinyML also facilitates micro-authentication, wherein the system verifies user identity repeatedly in short, transparent cycles without interrupting user experience. Examples include:

- Typing behavior analysis every few minutes,
- Voice or gesture recognition during active interaction,
- Facial micro-expression recognition during device use.

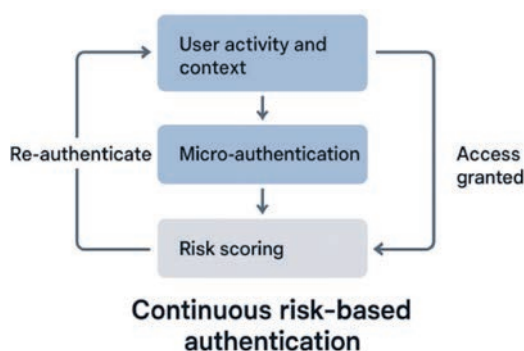


Figure 3.
Flow of continuous risk scoring or micro-authentication loop.

Threat vector	Example attack	Mitigation method	Covered section
Credential Theft	Phishing, password reuse	Ephemeral Tokens, Passwordless Login	Section 3
Biometric Spoofing	3D masks, voice replay	On-device model + Liveness Detection	Section 2
Model Tampering	OTA injection	Secure Boot + Signed Updates	Section 4
Model Extraction	Firmware reverse-engineer	TEEs, Obfuscation, Quantization	Section 4
Continuous Session Hijack	Device theft	Micro-authentication, Risk Scoring	Section 5
Data Privacy Violation	Cloud data leakage	Federated Learning, No raw data sent	Section 6

Table 2.
Threat mitigation techniques across TinyML identity layers.

This contextual authentication model ensures that access is tied to ongoing user presence and behavior not a one-time login event. Such mechanisms are critical in high-security environments like healthcare, finance, and critical infrastructure [23].

5.4 Role of EdgeAI in trust minimization

EdgeAI systems, equipped with more computational capabilities than TinyML, can run multi-modal models that combine various input signals (e.g., location, behavior, biometric) to generate a trust vector per user or device. This trust vector dynamically influences access policies, enabling fine-grained enforcement of least-privilege access control. Threat Mitigation Techniques Across TinyML Identity Layers are listed down in **Table 2** along with the section number where these topics are covered in this chapter. Additionally, EdgeAI supports federated learning and personalized risk modeling, allowing identity verification to become smarter and more adaptive over time without compromising privacy [14].

6. Federated identity and privacy

While on-device inference boosts privacy, long-term adaptability requires learning across users without compromising individual data. This section explores how federated learning extends TinyML’s capabilities into decentralized identity modeling. Traditional biometric and behavioral identity systems often rely on centralized cloud servers for training and updating machine learning models. This architecture poses significant privacy and security challenges, as raw personal data including sensitive biometric signals must be transmitted, stored, and processed outside of the user’s control. In contrast, federated learning, combined with TinyML, offers a privacy-preserving alternative for building robust, decentralized identity systems.

6.1 Federated learning for TinyML identity models

Federated learning allows multiple devices to collaboratively train or update a shared machine learning model without sharing raw data. Each device trains the

model locally using user-specific data (e.g., voice, gait, ECG), then sends only the updated model weights or gradients to a coordinating server. These updates are aggregated to improve a global model, which is redistributed to all devices [14, 25].

In the context of identity verification:

- A smartphone can fine-tune a TinyML voice model using the user's speech data,
- A smartwatch can adapt a heart rhythm model using ECG or PPG signals,
- A keyboard can update typing pattern models all without leaking sensitive data off-device.

This approach enables personalization and adaptability while preserving user data privacy, making it highly compatible with privacy laws such as GDPR and HIPAA (Health Insurance Portability and Accountability Act – USA) [14, 26].

6.2 Edge-centric identity learning across devices

TinyML and federated learning also support cross-device authentication ecosystems. A user's phone, watch, and smart speaker can each maintain a personalized identity model. When one device initiates authentication, others can validate the behavior or signal in parallel, contributing to a multi-source confidence score all without requiring a centralized data store [24].

This model of federated identity offers:

- Redundancy: If one biometric modality is unavailable (e.g., voice in noisy settings), another (e.g., gesture, heartbeat) can fill in.
- Increased robustness: Compromising one device does not expose the entire identity graph.
- On-device adaptation: Personalized models improve over time based on local context and interaction patterns.

Recent works have demonstrated the feasibility of federated TinyML training on microcontroller-class devices using platforms like TensorFlow Federated, Flower, and Edge Impulse, showing strong performance in use cases such as keyword spotting, anomaly detection, and behavioral verification [11, 25].

6.3 Privacy enhancements with differential privacy and encryption

To further reinforce data protection, federated learning implementations can incorporate:

- Differential privacy, which adds statistical noise to model updates before sharing, ensuring that individual data points cannot be reconstructed [14].
- Homomorphic encryption or secure multi-party computation, allowing model aggregation without decrypting intermediate values [26].

These mechanisms allow decentralized identity models to evolve over time while remaining provably private a critical requirement for health tech, finance, and national digital ID platforms. By embedding federated learning into TinyML-based authentication systems, we can create adaptive, privacy-first identity infrastructures that minimize cloud dependency, prevent mass surveillance, and provide users with greater control over their digital presence. To demonstrate the practical viability of the proposed concepts, next section presents real-world deployments and use cases that leverage EdgeAI and TinyML in identity verification.

7. Use cases and case studies

The integration of TinyML and EdgeAI into digital identity systems is not just theoretical it is actively reshaping the authentication landscape across consumer electronics, enterprise security, and critical infrastructure. This section presents several real-world use cases as shown in **Figure 4**, and case studies that demonstrate the viability and impact of these technologies,as listed in **Table 3** that highlights the Real-World Deployment Examples of EdgeAI Identity.

7.1 FaceID, windows hello, and secure enclaves in personal devices

One of the most widely adopted implementations of secure on-device identity verification is Apple’s FaceID, which performs 3D facial recognition using a local neural engine within the iPhone’s Secure Enclave. The biometric features are processed entirely on-device, with no raw data ever leaving the chip, and authentication tokens are cryptographically signed for secure interaction with online services.

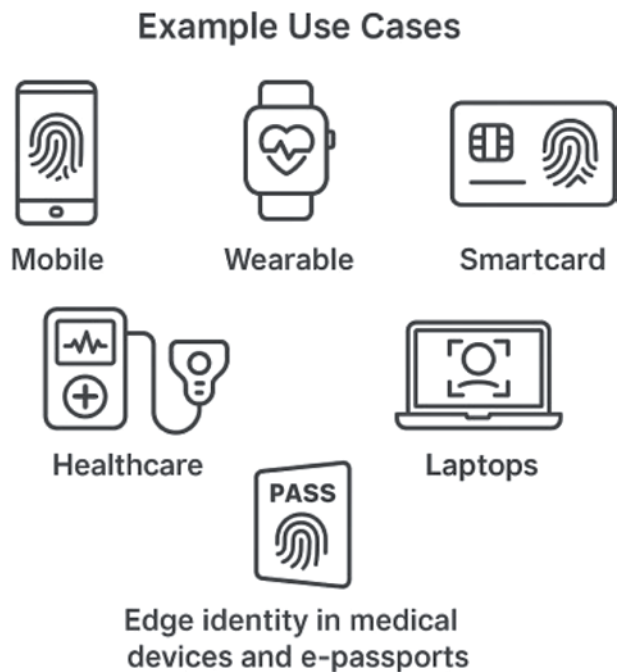


Figure 4.
Example use cases of authentication on edge devices.

System/ company	Modality	Model inference location	Uses TinyML	Privacy level	Reference
Apple FaceID	Face (3D)	Neural Engine (on-device)	Yes	High	[12]
Windows Hello	Face/ Fingerprint	TPM/Pluton Module	Yes	High	[27]
Google Gboard	Typing Behavior	On-device Federated	Yes	Very High	[14, 26]
Smart ECG Band	ECG	Microcontroller (MCU)	Yes	Very High	[21]
E-passport	Face/ Fingerprint	Embedded chip + verifier	Partially	High	[28]

Table 3.
Real-world deployment examples of EdgeAI identity.

Similarly, Microsoft’s Windows Hello provides facial recognition and fingerprint authentication on laptops and tablets using infrared cameras or fingerprint sensors. The biometric processing occurs in a TPM (Trusted Platform Module) or Pluton Security Processor, which operates in a similar fashion to a TEE [27]. These identity systems leverage local TinyML-style models to perform inference securely and in real-time:

- Face matching occurs in less than 100 ms on-device.
- Fingerprint matching is handled by hardware modules with embedded ML algorithms.
- No passwords are needed credentials are replaced by cryptographic tokens.

These technologies align closely with WebAuthn and FIDO2 standards, enabling passwordless login to web platforms and enterprise services across mobile, desktop, and cloud environments. As TinyML tools continue to evolve, increasingly personalized and lightweight identity models will be integrated into endpoint devices for cross-platform access with local authentication guarantees.

7.2 Federated learning in Gboard and smart keyboards

Google’s Gboard keyboard is a practical example of federated learning in identity-related personalization. While not a biometric system per se, it uses local learning from users’ typing patterns to suggest better predictions without sending text data to cloud servers. The underlying architecture mirrors privacy-centric identity systems: on-device TinyML models learn context-aware behavior, update locally, and only send encrypted gradients back to the cloud [14, 26]. This model can be easily adapted to behavioral identity use cases such as keystroke biometrics in enterprise or high-security authentication systems [19].

7.3 Continuous authentication in wearables

Modern wearables such as smartwatches, fitness bands, and smart rings offer a rich source of physiological and behavioral data. These devices can perform

continuous passive authentication using PPG, ECG, motion sensors, or skin conductance. A TinyML model trained on such biosignals can determine whether the current user is still the authorized user without interrupting the session or requiring active input [3, 20, 29].

Startups and research groups have demonstrated prototype systems that identify users based on gait, typing behavior, or even heartbeat morphology, processed in real time on devices like the ESP32, nRF52, or Ambiq Apollo3 platforms [11, 20]. Beyond consumer devices, these principles are now being adopted in infrastructure-critical applications such as smart access systems and national ID frameworks.

7.4 AI-powered smart access systems

Edge-based identity systems are also being deployed in physical access control scenarios:

- Smart door locks use voice or facial recognition verified locally via TinyML to authenticate users.
- Industrial badge systems integrate gesture or location-based authentication using embedded ML.
- Healthcare terminals use federated identity to confirm patient or provider access without relying on shared credentials.

These systems are particularly effective in offline environments, such as disaster zones, military outposts, or rural healthcare settings, where cloud access is limited or prohibited.

7.5 Privacy-centric national ID systems (emerging)

Governments and public agencies are exploring decentralized digital ID frameworks that prioritize privacy, mobility, and resilience. Although still early-stage, pilot initiatives are experimenting with TinyML-powered smartcards that store biometric verification models (e.g., fingerprint, iris, face) and execute inference locally [5]. Combined with federated learning, these systems could eliminate centralized biometric databases long seen as privacy risks and offer citizens greater control over their identity.

7.6 Edge identity in medical devices and e-passport infrastructure

7.6.1 Medical devices and patient identity verification

In healthcare environments, especially those involving telemedicine, portable diagnostics, or emergency response, verifying a patient's identity quickly and securely is crucial. Traditional identity checks are prone to manual errors, card loss, and impersonation. Recent research and pilot deployments have shown that EdgeAI and TinyML-enabled medical devices can:

- Authenticate patients using on-device fingerprint or ECG signature verification,

- Tie readings (e.g., insulin dosage, vital monitoring) to a verified patient profile, stored only locally,
- Ensure tamper-proof data logging for compliance with HIPAA or GDPR.

For instance, an edge diagnostic device might use a TinyML model to recognize a user's heart rhythm pattern from ECG input before administering medication. This prevents unauthorized access or patient misidentification, even in low-connectivity environments [20, 21]. E-passports (electronic passports) already include embedded RFID chips that store facial and fingerprint templates. Increasingly, on-device biometric verification modules are being explored to:

- Match the passport holder's live biometric (e.g., via face camera) directly within the document,
- Enable offline border verification without needing centralized lookups,
- Detect spoofing or tampering attempts at the point of control using embedded anomaly detection [28].

Several national security agencies are evaluating contactless, embedded ML models to verify passport holders with high assurance in constrained hardware very much aligned with TinyML principles. These systems can also integrate with mobile identity wallets and decentralized digital ID systems, where ephemeral tokens are generated per session, avoiding data reuse or exposure.

8. Ethical, legal, and regulatory considerations

Despite the technical advancements, real-world deployment of biometric and behavioral identity systems must address significant ethical and legal challenges. This section reviews these responsibilities in the context of EdgeAI and TinyML. The growing use of biometric and behavioral identity verification powered by TinyML and EdgeAI offers clear advantages in terms of privacy, performance, and decentralization. However, these technologies also raise critical ethical, legal, and regulatory concerns that must be addressed before widespread deployment especially in contexts where identity data can impact access to essential services or civil rights. A quick illustration of common ethical risk zones and their possible protections is listed in **Figure 5**.

8.1 Data privacy and consent

Biometric signals such as fingerprints, facial features, heart rhythms (ECG/PPG), and typing behavior are uniquely identifying, immutable, and potentially revealing of sensitive health or demographic traits. While TinyML offers enhanced privacy by processing data locally, users must still be informed about:

- What is being collected,
- How it is used and stored,

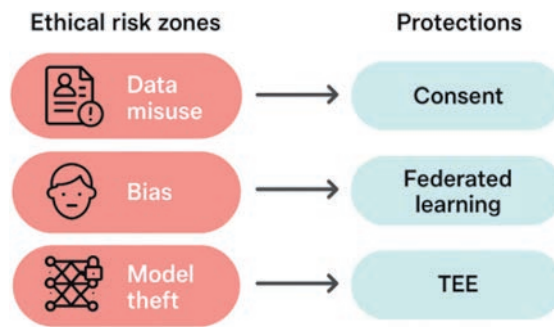


Figure 5.
Ethical risk zones vs. protections.

- Whether the model is personalized, and
- How they can opt out or revoke consent.

Regulatory frameworks such as the GDPR and California Consumer Privacy Act (CCPA) require explicit, informed consent for the collection and use of biometric identifiers [26, 30]. In decentralized systems, consent mechanisms should be device-specific and allow fine-grained control over biometric usage across applications.

8.2 Data ownership and control

Unlike passwords, biometrics cannot be reset. Once leaked or misused, a user's biometric profile may be permanently compromised. Therefore, user-centric data ownership is critical. Federated TinyML models reduce centralized risk but introduce new challenges around:

- Who owns the locally trained model,
- How updates are audited or revoked,
- What happens when devices are resold, shared, or stolen.

Designing identity systems with local model explainability, audit trails, and secure deletion capabilities is essential to maintaining transparency and control.

8.3 Algorithmic bias and fairness

Biometric systems are vulnerable to demographic bias, particularly when models are trained on limited datasets. This can lead to disparate outcomes for:

- Minority ethnic groups (in facial recognition),
- Women and older adults (in voice or gait recognition),
- Individuals with medical conditions (in heart rate variability or motion-based inference).

TinyML models trained locally may mitigate some of this by adapting to the user over time, but the risk of systemic bias from pre-trained global models still exists. Regulatory scrutiny (e.g., from the EU AI Act or NIST Fairness standards) is increasing on such algorithmic decisions [23, 31]. As we move into regulated domains like healthcare and border security, the compliance landscape becomes even more complex and demanding.

8.4 Legal compliance in regulated sectors

In fields such as healthcare, finance, and border control, legal requirements go beyond privacy to include:

- Auditability of authentication events,
- Tamper-proof logging for legal admissibility,
- Data minimization (collecting only what's strictly needed),
- Encryption at rest and in transit, even in federated or offline modes.

In healthcare, for example, compliance with HIPAA and ISO/IEEE 11073 is required for patient identity systems using biometric wearables [21]. In border security, ICAO (International Civil Aviation Organization) standards govern biometric e-passport validation processes [28].

8.5 Security transparency and trust

With increasing attacks on AI pipelines and edge devices, ethical system design must prioritize:

- Verifiability of model updates (signed and traceable),
- Public audit logs for identity verification decisions,
- Tamper-evident models and trust anchors at the hardware level.

Developers and manufacturers must be prepared to disclose security practices and engage third-party audits to establish trust with users and regulators alike.

9. Future outlook

Looking ahead, the convergence of EdgeAI, decentralized identity, and privacy-first design principles is poised to reshape the very fabric of online authentication. This section outlines the potential trajectories and innovations expected over the next decade. The integration of TinyML and EdgeAI into digital identity systems marks a significant paradigm shift from static, cloud-dependent authentication models to decentralized, privacy-preserving, and context-aware systems. As edge devices become increasingly intelligent, identity verification will evolve from a one-time gate-keeper function to a continuous, adaptive layer of personal and societal interaction.

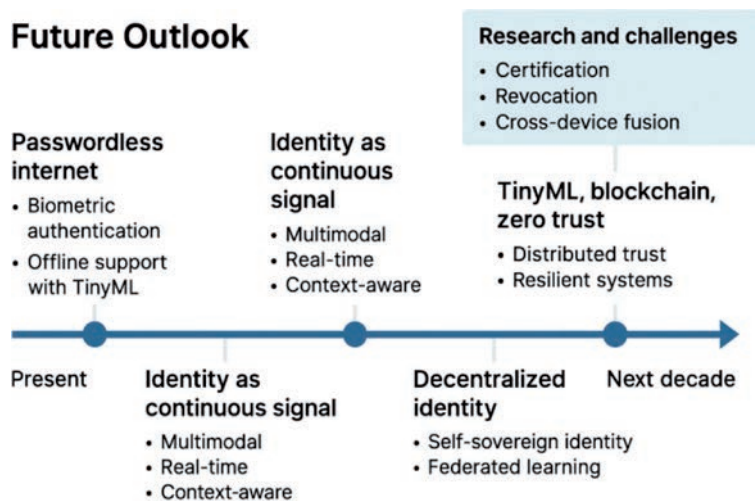


Figure 6.
Future Outlook visual roadmap representation.

The predicted future advancements of Edge AI and TinyML implementations in terms of online identity and biometrics is illustrated in **Figure 6**.

9.1 Passwordless internet and ubiquitous identity

The widespread adoption of passwordless login, driven by standards such as FIDO2 and WebAuthn, is accelerating. Over the next decade, users will increasingly rely on biometric-backed ephemeral tokens generated by their personal devices be it a phone, smartwatch, laptop, or smartcard to access services across domains, including:

- Banking,
- Government portals,
- Enterprise systems,
- Smart home ecosystems,
- Healthcare records.

TinyML enables this vision even in offline, resource-constrained, and remote environments, where traditional cloud services are inaccessible or impractical.

9.2 Identity as a continuous, multimodal signal

Static identity checks will give way to multi-modal, real-time behavioral signatures, verified through a combination of:

- Physiological signals (e.g., ECG, PPG),

- Behavioral patterns (e.g., typing, walking, speaking),
- Environmental context (e.g., location, proximity to other devices).

TinyML will allow this continuous identity stream to be analyzed locally, creating identity-aware devices that constantly validate presence and context minimizing fraud risk and enhancing user experience without invasive prompts.

9.3 Decentralized identity networks

The future of online identity lies in self-sovereign identity (SSI) and decentralized ID ecosystems. In such architectures:

- Users control their digital credentials across platforms,
- Identity wallets are managed locally and interoperably,
- Federated learning ensures models adapt privately over time.

TinyML plays a key role in making SSI accessible on low-power devices, bridging the gap between high-security standards and real-world usability especially in developing regions or vulnerable populations.

9.4 Merging TinyML with blockchain and zero trust

Emerging innovations will combine TinyML, ZTA, and blockchain for auditability, transparency, and secure distributed trust. For example:

- Local biometric verification triggers a smart contract on a blockchain to log the event,
- Federated identity scores are hashed and recorded as tamper-evident proof,
- IoT ecosystems enforce identity-based access policies without centralized brokers.

Such systems promise resilient and censorship-resistant identity layers, critical for humanitarian access, crisis response, and digital democracy.

9.5 Research frontiers and open challenges

Despite these promising trends, several open questions remain:

- How do we evaluate and certify TinyML-based identity systems?
- How do we handle revocation of local models in case of compromise?
- What are the best practices for cross-device identity fusion?
- How can we ensure algorithmic fairness and explainability in low-resource models?

Addressing these questions will require cross-disciplinary collaboration between engineers, ethicists, regulators, and civil society.

In summary, TinyML and EdgeAI are laying the groundwork for a future where identity is secure, portable, adaptive, and private by design. As we transition from centralized credentials to intelligent, self-owned authentication mechanisms, the next generation of online identity will be embedded, invisible, and individually empowered.

10. Conclusion

Bringing together the technical, ethical, and practical dimensions discussed in the previous sections, this conclusion summarizes the key insights and defines the road ahead for privacy-preserving identity systems. The digital identity landscape is undergoing a fundamental transformation. Traditional cloud-based authentication systems are being gradually replaced or augmented by TinyML-powered, on-device solutions that offer significant improvements in privacy, security, latency, and adaptability. This evolution is not merely a technological shift it redefines how individuals interact with digital systems, granting them more control over their identity and personal data.

Throughout this chapter, we explored the core building blocks of TinyML-based identity verification:

- The biometric signal acquisition from physiological sources like ECG and PPG,
- Local inference and decision-making via resource-efficient models,
- Integration with Zero Trust Architecture to support continuous, risk-based authentication,
- Privacy-enhancing mechanisms such as federated learning and secure boot pipelines to protect both models and user data,
- Deployment in real-world use cases across mobile, wearable, smartcard, laptop, medical, and passport systems.

These developments are not occurring in isolation. They are part of a broader convergence of trends in edge computing, secure hardware, privacy-aware AI, and decentralized identity protocols. As our future systems become more context-aware, user-centric, and autonomous, the role of on-device intelligence in authentication will only grow. However, challenges remain. Ethical questions around data consent, model bias, and identity ownership require continued scrutiny. Regulatory frameworks must evolve in parallel with technical advancements. And developers must prioritize transparency and inclusivity as these technologies scale globally.

In closing, TinyML and EdgeAI represents not only a solution for efficient computation it represents a new model of trust. One where identity is verified at the edge, owned by the user, and constantly reinforced by behavior and context rather than static credentials. Embracing this paradigm will lead us toward a future where authentication is seamless, secure, and human-centered by design. Together, these innovations set the foundation for the next generation of trustable, adaptive, and secure online identity systems designed for the edge, and centered on the user.

Author details

Yogeswar Reddy Thota* and Tooraj Nikoubin
Department of Electrical and Computer Engineering, University of Texas at Dallas,
Richardson, TX, United States of America

*Address all correspondence to: yogeswarreddy.thota@utdallas.edu

IntechOpen

© 2025 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Nguyen Ba MH, Bennett J, Gallagher M, Bhunia S. A case study of credential stuffing attack: Canva data breach. In: 2021 International Conference on Computational Science and Computational Intelligence (CSCI). Las Vegas, NV, USA; 2021. pp. 735-740. DOI: 10.1109/CSCI54926.2021.00187
- [2] Das A, Borisov N. The Tangled Web of Password Reuse. NDSS Symposium. 2014. Available from: <https://www.ndss-symposium.org/ndss2014/tangled-web-password-reuse>
- [3] Wazid M, Das AK, Odelu V, Kumar N, Conti M, Jo M. Design of secure user authenticated key management protocol for generic IoT networks. IEEE Internet of Things Journal. Feb 2018;5(1):269-282. DOI: 10.1109/JIOT.2017.2780232
- [4] Sasi T, Lashkari AH, Lu R, Xiong P, Iqbal S. Comprehensive survey on IoT attacks: Taxonomy, detection mechanisms and challenges. Journal of Information and Intelligence. 2024;2(6):455-513. ISSN 2949-7159. DOI: 10.1016/j.jiixd.2023.12.001
- [5] Jain AK, Nandakumar K, Nagar A. Biometric template security. EURASIP Journal on Advances in Signal Processing. 2008;2008:113. DOI: 10.1155/2008/579416
- [6] Alrawili R, AlQahtani AAS, Khan MK. Comprehensive survey: Biometric user authentication application, evaluation, and discussion. Computers and Electrical Engineering. 2024;119(Part A):109485. DOI: 10.1016/j.compeleceng.2024.109485. ISSN 0045-7906
- [7] Elordi U et al. Designing automated deployment strategies of face recognition solutions in heterogeneous IoT platforms. Information. 2021;12(12):532. DOI: 10.3390/info12120532
- [8] Preukschat & Reed. Self-Sovereign Identity. Manning Publications; 2021. Available from: <https://www.manning.com/books/self-sovereign-identity>
- [9] Sundararajan A, Sarwat AI, Pons A. A survey on modality characteristics, performance evaluation metrics, and security for traditional and wearable biometric systems. ACM Computing Surveys. 2019;52(2):36, Article 39 (March 2020). DOI: 10.1145/3309550
- [10] Hahn VK, Marcel S. Biometric template protection for neural-network-based face recognition systems: A survey of methods and evaluation techniques. IEEE Transactions on Information Forensics and Security. 2023;18:639-666. DOI: 10.1109/TIFS.2022.3228494
- [11] Warden P, Situnayake D. Tinyml: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers. O'Reilly Media; 2019
- [12] Zhou Z, Chen X, Li E, Zeng L, Luo K, Zhang J. Edge intelligence: Paving the last mile of artificial intelligence with edge computing. Proceedings of the IEEE. 2019;107(8):1738-1762. DOI: 10.1109/JPROC.2019.2918951
- [13] Zerraza I, Seghir Z, Hemam M. An efficient lightweight authentication and access control for IoT edge devices. International Journal of Safety and Security Engineering. 2024;14:807-813. DOI: 10.18280/ijss.140313
- [14] McMahan HB, Ramage D, Talwar K, Zhang L. Learning differentially private recurrent language models. In: International Conference on Learning

- Representations (ICLR). ArXiv. 2018. Available from: <https://arxiv.org/abs/1710.06963>
- [15] Liu X, Xie L, Wang Y, Zou J, Xiong J, Ying Z, et al. Privacy and security issues in deep learning: A survey. IEEE Access. 2020;1-1. DOI: 10.1109/ACCESS.2020.3045078
- [16] Papernot N, McDaniel P, Goodfellow I, Jha S, Celik ZB, Swami A. Practical black-box attacks against machine learning. In: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security (ASIA CCS '17). New York, NY, USA: Association for Computing Machinery; 2017. pp. 506-519. DOI: 10.1145/3052973.3053009
- [17] Pang R, Adam H, Le Q, Sandler M, Chen B, Wang W, et al. Searching for MobileNetV3. 2019. pp. 1314-1324. DOI: 10.1109/ICCV.2019.00140
- [18] Zhang Y, Suda N, Lai L, Chandra V. Hello Edge: Keyword Spotting on Microcontrollers. 2017. DOI: 10.48550/arXiv.1711.07128
- [19] Banerjee S, Woodard DL. Biometric authentication and identification using keystroke dynamics: A survey. Journal of Pattern Recognition Research. 2012;7(1):116-139
- [20] Thota YR, Nikoubin T. TinyML for ECG biometrics on resource constrained devices. In: 2024 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), Knoxville, TN, USA. 2024. pp. 521-526. DOI: 10.1109/ISVLSI61997.2024.00100
- [21] Wang G, Shanker S, Nag A, Lian Y, John D. ECG biometric authentication using self-supervised learning for IoT edge sensors. IEEE Journal of Biomedical and Health Informatics. 2024;28(11):6606-6618. DOI: 10.1109/JBHI.2024.3455803
- [22] Uchida Y, Nagai Y, Sakazawa S, Satoh S. Embedding watermarks into deep neural networks. In: Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval (ICMR '17). New York, NY, USA: Association for Computing Machinery; 2017. pp. 269-277. DOI: 10.1145/3078971.3078974
- [23] Mohseni-Ejyeh A. Zero trust real time lightweight access control protocol for mobile cloud-based IoT sensors. ArXiv. 2023. Available from: <https://arxiv.org/abs/2309.01293>
- [24] Liu Y, Wang Y, Chen H, Liu A. Toward zero trust IoT security: Lightweight device authentication in edge computing. IEEE Internet of Things Journal. 2021;8(14):11365-11378. DOI: 10.1109/JIOT.2021.3053729
- [25] Li T, Sahu AK, Talwalkar A, Smith V. Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine. 2020;37(3):50-60. DOI: 10.1109/MSP.2020.2975749
- [26] Bonawitz K, Eichner H, Grieskamp W, et al. Towards federated learning at scale: System design. In: Proceedings of the 2nd SysML Conference. 2019. Available from: <https://arxiv.org/abs/1902.01046>
- [27] Sun Q, Wu J, Yu W. BioShare: An open framework for trusted biometric authentication under user control. Applied Sciences. 2022;12:10782. DOI: 10.3390/app122110782
- [28] Choudhury Z. Encryption and encoding of facial images into quick response and high capacity color 2d code for biometric passport security system. 2022. DOI: 10.48550/arXiv.2203.15738

- [29] Thota YR, Nixon JS, Chandran B, Nikoubin T. TinyML based biometric authentication using PPG signals for edge devices. In: Proceedings of the Great Lakes Symposium on VLSI 2025 (GLSVLSI'25). New York, NY, USA: Association for Computing Machinery; 2025. pp. 860-865. DOI: 10.1145/3716368.3735271
- [30] Voigt P, Von dem Bussche A. The EU General Data Protection Regulation (GDPR): A Practical Guide. Springer International Publishing; 2017. DOI: 10.1007/978-3-319-57959-7
- [31] Grother P, Ngan M, Hanaoka K. Face Recognition Vendor Test Part 3: Demographic Effects. NIST Interagency/Internal Report (NISTIR). Gaithersburg, MD: National Institute of Standards and Technology; 2019. [Online]. DOI: 10.6028/NIST.IR.8280

Artificial Intelligence, Trust, and the Architecture of Online Identity: Security Implications in an Algorithmic Era

Livinus Obiora Nweke

Abstract

The increasing integration of artificial intelligence (AI) into digital identity systems is reshaping how online identities are formed, authenticated, and managed. As algorithmic decision-making becomes more deeply embedded in identity infrastructures, from facial recognition and behavioral analytics to AI-powered verification platforms, questions of trust, accountability, and security grow increasingly urgent. This chapter examines the evolving architecture of online identity in the context of AI-driven technologies, with a focus on their implications for infrastructure security, privacy, and digital assurance. Through a multidisciplinary lens, the chapter explores how AI alters traditional models of trust and challenges the assumptions underpinning user consent, transparency, and data protection. It investigates the role of algorithmic bias, the opacity of machine learning models, and the potential for manipulation *via* synthetic identity and deepfake technologies. In response, the chapter proposes a security-centric framework for the responsible integration of AI into identity systems, drawing on principles from risk management, information security, and privacy by design. This contribution offers practical insights and strategic recommendations for policymakers, system architects, and security professionals tasked with safeguarding online identities in an increasingly algorithmic world.

Keywords: artificial intelligence, trust, online identity, identity infrastructure security, privacy, digital assurance

1. Introduction

The concept of identity has undergone a profound transformation in today's increasingly digital society. Once tied to physical documents and in-person authentication, identity is now often established, verified, and managed online. This evolution, accelerated by the rise of digital platforms and remote services, has redefined how individuals prove who they are and how institutions recognize and authorize them [1–3]. Amid this shift, online identity has emerged as a foundational

component of digital interactions, underpinning services ranging from banking and healthcare to commerce and governance [4].

At the same time, artificial intelligence (AI) has emerged as a defining force in shaping digital systems. AI technologies, particularly machine learning (ML), natural language processing, and computer vision, are now increasingly embedded in identity infrastructures to enhance fraud detection, streamline onboarding, enable biometric recognition, and automate access control [5–7]. These applications promise improved efficiency, scalability, and adaptability, which make them attractive to both private and public sector actors seeking to secure and personalize identity services.

However, the convergence of AI and online identity introduces new complexities. This is because, as AI becomes the intermediary in identity-related decisions, such as verifying a biometric sample, analyzing user behavior, or flagging anomalies in real time, it reshapes the architecture of trust [8]. For example, in traditional systems, trust was mediated by human institutions and regulated frameworks [9]. In contrast, recent developments in digital identity systems have now extended trust to data-driven models that may operate opaquely and autonomously [10]. This shift raises pressing questions:

- How does the integration of AI alter the security posture of digital identity systems, particularly in terms of vulnerabilities introduced by algorithmic components?
- What are the privacy implications when identity systems rely on extensive data collection, behavioral profiling, or real-time inference?
- In what ways does AI reinforce or undermine trust in identity systems, especially when outcomes are not easily explainable?
- What governance models and technical safeguards are needed to ensure that AI-enhanced identity systems are transparent, accountable, and resilient?

Therefore, this chapter responds to the urgent need for an interdisciplinary and security-focused examination of AI's impact on the architecture, trust dynamics, and security of online identity. It draws on insights from information security, privacy engineering, digital assurance, and risk governance to explore how AI is reshaping the boundaries, trust mechanisms, and vulnerabilities of modern identity systems. Through the analysis of the interplay between AI and online identity, this chapter aims to provide scholars, practitioners, and policymakers with a lens through which to assess the security and trustworthiness of emerging identity paradigms.

The rest of the chapter is structured as follows. Section 2 explores the foundational models of online identity, detailing how AI is being integrated into centralized, federated, and decentralized architectures. Section 3 examines the impact of AI on trust in identity systems and the implications of algorithmic opacity. Section 4 assesses the security implications of AI, including attack surfaces introduced by machine learning, adversarial threats, and system-level vulnerabilities. Section 5 discusses privacy risks and compliance concerns, particularly in relation to profiling, consent, and data minimization. Section 6 proposes a security-centric integration framework for AI in identity systems, emphasizing design principles and governance tools. Section 7 presents case studies across sectors to illustrate practical applications and challenges. Section 8 highlights future research directions, and Section 9 concludes the chapter.

2. The architecture of online identity: Foundations and evolution

This section presents the foundational models and evolution of online identity. It also highlights the technological enablers of modern identity systems and the impact of AI across identity architectures.

2.1 Identity system architectures

Online identity systems are foundational to digital interaction as they enable individuals and entities to authenticate themselves, access resources, and establish trust [4]. Over time, the architecture of these systems has evolved from centralized structures controlled by single authorities to more federated and recently decentralized/self-sovereign models (which distribute control and empower users). Each model is defined by how identity data is stored, who controls access, and how verification occurs.

Centralized identity architectures: Rely on a single trusted authority, often a government agency, enterprise, or service provider, that stores and verifies user credentials [11]. Examples include state-issued ID databases or enterprise directory services. While these systems offer administrative control and policy uniformity, they present single points of failure. This is because a compromise of the central authority could expose the identity data of millions.

Federated identity architectures: Enable users to authenticate across multiple services using a single set of credentials managed by an identity provider (IdP) [12]. For example, signing into third-party applications using Google or Microsoft accounts. Federated identity improves convenience and interoperability but can introduce dependency risks on the IdP and privacy trade-offs, as usage data may be aggregated across services. While federated architectures can reduce the need for repetitive credential management, they centralize trust in IdPs, thereby increasing exposure to systemic failures or consent fatigue. These weaknesses can be mitigated by AI-driven risk scoring, but they may also raise concerns about fairness and explainability when opaque decision models are used.

Decentralized and self-sovereign identity (SSI) architectures: Allow users to control their identity data using cryptographic keys stored in digital wallets [13]. Verification is enabled through verifiable credentials issued by trusted entities but held by users themselves. This model seeks to restore data sovereignty and reduce reliance on intermediaries. SSI is supported by decentralized identifiers (DIDs) and often built atop blockchain or distributed ledger technologies.

2.2 Key components of identity systems

Across all models, several core components enable the functioning of identity systems:

- *Identifiers:* These are unique strings (e.g., usernames, email addresses, and DIDs) used to reference entities within a system.
- *Credentials:* Digital artifacts (e.g., passwords, tokens, biometric templates, and certificates) that prove identity claims.
- *Verification processes:* Methods of authenticating credentials and validating claims, which may be static (e.g., password matching) or dynamic (e.g., behavioral analysis).

Identity verification also involves a lifecycle: registration, authentication, authorization, and auditing; each of which can be enhanced or undermined by the technologies employed.

2.3 Technological enablers of modern identity systems

Modern identity systems have also been transformed by an ecosystem of enabling technologies:

- *Cloud computing*: Enables scalable identity services, such as Identity-as-a-Service (IDaaS) [14], which reduces the burden of on-premise identity infrastructure.
- *Mobile devices*: Smartphones now serve as primary platforms for digital identity as they offer multifactor authentication (MFA), biometric sensors, and secure enclaves [15].
- *Biometrics*: Fingerprint, facial recognition, voice, and iris scans are increasingly used for identity proofing, especially in financial services and border control. While biometrics enhance convenience, they also raise immutability and consent challenges [16].
- *Blockchain*: Distributed ledgers provide the foundation for tamper-evident, trustless identity systems. DIDs and verifiable credentials rely on blockchain for public key registration and credential anchoring, especially in SSI ecosystems [13].

2.4 The role of AI across identity architectures

AI is being deployed across different identity architectures to improve service delivery. This allows AI to serve as a cross-cutting enabler that influences each layer of the identity stack, from registration and verification to fraud detection and access management.

AI in centralized identity architectures: In centralized identity architectures, AI is used primarily for pattern recognition, anomaly detection, and biometric verification. For instance, ML algorithms can analyze transaction patterns to flag potential identity theft or account takeover [17]. AI can also automate identity-proofing processes by extracting and validating data from scanned ID documents using computer vision and optical character recognition (OCR) [18]. However, reliance on opaque models can reduce explainability and accountability.

AI in federated identity architectures: Federated identity architectures leverage AI to personalize user experience and perform risk-based authentication. For instance, AI can assess behavioral signals such as typing patterns, location data, and device signatures to determine whether a login attempt should be challenged with additional factors [19]. These systems can dynamically adapt authentication requirements based on the calculated risk level, which will help enhance usability while preserving security. Although adaptive authentication improves security and user experience, it introduces privacy risks by continuously collecting behavioral and contextual data. Over-reliance on opaque models can also create biases and false rejections, disproportionately impacting marginalized demographics [20].

AI in decentralized and SSI architectures: While SSI architectures emphasize decentralization and user control, AI plays a growing role in user-agent decision

Architecture	Trust anchor	User control	AI integration	Risks introduced
Centralized	Single authority	Low	Biometric verification, fraud detection	Limited transparency, adversarial attacks
Federated	Trusted identity providers	Moderate	Adaptive authentication, behavioral analytics	Cross-domain profiling, consent ambiguity, bias training
Decentralized/SSI	User and verifiers	High	Smart contracts, reputation analysis	Usability, interoperability, and governance gaps

Table 1.
Comparative analysis of identity architectures and AI integration points.

support, such as recommending which credentials to present or flagging suspicious requests [21]. AI can also enhance biometric privacy in decentralized systems by enabling homomorphic encryption or federated learning to train models without exposing raw data [22]. However, integrating AI with decentralized infrastructure presents design challenges, especially around trust and governance.

Emerging risks and benefits: As shown in the comparative analysis of identity architectures and AI integration points in **Table 1**, AI can improve speed and efficiency across identity architectures, but it also introduces new attack surfaces, such as adversarial inputs that trick facial recognition systems [23]. Moreover, biased training data can lead to discriminatory outcomes in automated decision-making, which will further reinforce social inequities [24, 25]. Therefore, approaches to ensure transparency, auditability, and fairness in AI-powered identity systems are critical and imperative to design.

3. Trust and AI in digital identity systems

This section explores how AI reshapes the trust landscape in digital identity ecosystems, examines the trade-offs between explainability and automation, and considers the social and psychological dimensions of trusting AI-driven identity technologies.

3.1 The algorithmic shift in trust models

In traditional identity systems, trust was anchored in centralized authorities, such as governments or financial institutions. However, with the proliferation of AI-powered technologies, such as biometric authentication, intelligent fraud detection, and adaptive access controls, the concept of trust has increasingly shifted to data-driven decision systems [8]. For example, ML algorithms now play a central role in authenticating users based on behavioral biometrics, such as keystroke dynamics [26], often without the user's direct awareness.

This shift challenges the foundational assumptions of trust in identity management. This is because, unlike traditional identity systems, AI systems derive their trustworthiness from perceived technical robustness, safety, consistency, and security [27]. As such, users and stakeholders must decide whether to trust the AI's judgment, often without understanding its internal logic or having clear recourse in the case of error.

3.2 Trust challenges in opaque (black-box) AI algorithms

A central challenge in AI-integrated identity systems is opacity. Many high-performing models, particularly deep learning architectures, function as “black boxes,” producing outcomes without readily interpretable reasoning paths. This is particularly problematic in contexts such as identity verification or fraud prevention, where decisions may have significant personal, legal, or financial implications.

For instance, an AI system may deny access to a user based on subtle biometric mismatches detected by convolutional neural networks, without offering a human-readable rationale. This opacity can erode user confidence, raise concerns of bias and unfairness, and complicate regulatory compliance, especially with frameworks like the European Union’s General Data Protection Regulation (GDPR) [28], requiring the implementation of suitable safeguards against automated decision-making [29].

Moreover, opaque AI can exacerbate issues of algorithmic bias. If training data reflects societal inequities or lacks diversity, the AI may systematically disadvantage certain groups, which undermines trust in both the system and the institutions deploying it. Notably, studies have shown that facial recognition algorithms often perform poorly on individuals with darker skin tones, introducing disproportionate access errors [20, 30, 31]. Also, the evaluations of facial recognition systems show significant demographic disparities, with false positive rates generally higher for Black, East Asian, and American Indian individuals compared to white individuals and higher for women compared to men [32]. These disparities highlight the risks of embedding opaque AI systems into identity verification pipelines.

3.3 Explainability vs. automation: Balancing performance and transparency

The tension between explainability and automation presents a significant design dilemma. On one hand, highly automated systems offer speed, scalability, and predictive power, which are key advantages in managing large-scale identity ecosystems. On the other hand, explainability fosters accountability, user trust, and legal defensibility.

To address this, the field of Explainable AI (XAI) has emerged, offering tools such as Shapley Additive Explanations (SHAP) [33] or Local Interpretable Model-agnostic Explanations (LIME) [34] to approximate explanations for otherwise opaque models. While promising, these methods often provide post-hoc rationalizations rather than true transparency, and their effectiveness in identity-sensitive contexts remains contested.

An emerging strategy involves hybrid architectures, where AI handles routine decisions while deferring ambiguous or high-risk cases to human operators. This “human-in-the-loop” (HITL) approach helps balance the efficiency of automation with the ethical and practical demands of explainability and user agency. For example, biometric authentication platforms may automatically approve common access requests but flag anomalies for manual review, thereby preserving user trust while minimizing friction.

3.4 Social and psychological dimensions: User trust and digital reputation

Beyond technical design, trust in AI-enabled identity systems also hinges on social and psychological factors. Research shows that users tend to anthropomorphize intelligent systems, attributing human-like qualities to algorithmic agents, especially

when systems provide feedback or simulate empathy [35]. However, this effect can be fragile as any perceived failure, lack of clarity, or misalignment with user expectations can quickly erode trust.

Digital reputation systems, such as those embedded in e-commerce platforms like Airbnb [36], can further complicate trust dynamics. These systems often use AI to score user behavior, assess credibility, or detect fraud. While useful, they can also reinforce biases, limit second chances, and create “digital shadows” that persist across platforms.

In the identity context, such reputational markers may become proxies for trustworthiness, affecting access to services or opportunities. This raises profound ethical questions: Should an AI be allowed to infer identity legitimacy from social behavior? Who governs the algorithms that shape digital reputations? How can users contest or correct their digital profiles?

To foster sustainable trust, developers and policymakers must address these socio-technical complexities. Strategies include building participatory design frameworks, establishing independent auditing bodies, and enhancing algorithmic literacy among users.

4. Security implications of AI-augmented identity systems

This section examines emerging vulnerabilities and AI-driven threats that affect digital identity ecosystems, while also exploring how the introduction of AI impacts the security of the supporting infrastructure.

4.1 Vulnerabilities in AI-enhanced identity systems

AI systems, particularly those based on ML, are susceptible to several classes of attacks that undermine the integrity of identity verification processes. Some of these attacks include:

- *Adversarial attacks:* These are attacks that use adversarial examples that are designed to mislead ML models [37]. In the context of biometric authentication (e.g., facial or fingerprint recognition), small perturbations that are imperceptible to humans can cause AI models to misclassify or misidentify individuals. For instance, attackers can wear specially designed glasses or apply facial patterns (known as adversarial patches) to bypass facial recognition systems [23].
- *Model poisoning:* In model poisoning attacks, adversaries can inject manipulated data during the training phase of an AI model [38]. This may cause the model to behave erroneously under specific conditions, allowing attackers to create “backdoors” that grant unauthorized access. In federated identity systems where decentralized models may be collaboratively trained, this vulnerability becomes particularly acute [39].
- *Data leakage and inference risks:* Training AI models on sensitive identity data can introduce risks of unintended data leakage. Malicious actors can exploit model outputs to infer private attributes of users, reconstruct training data, or even deanonymize individuals [40]. These risks are amplified in scenarios where models are deployed across shared or cloud-based infrastructures.

4.2 AI-driven threats to identity integrity

While AI can serve as a powerful tool for securing identity systems, it also empowers adversaries with unprecedented capabilities to perpetrate sophisticated attacks. Common AI-driven threats to identity integrity include:

- *Deepfake identity fraud*: AI-generated synthetic media, known as “deepfakes,” are increasingly used to impersonate individuals in video and audio communications. These tools pose a severe threat to remote identity verification services. Fraudsters can exploit deepfake videos to defeat liveness detection protocols and fool biometric authentication systems [41].
- *Synthetic identity generation*: AI models trained on large datasets can fabricate plausible, non-existent identities, known as synthetic identities, that evade traditional detection techniques [42]. Such identities often blend real and fake data, making them particularly difficult to identify. Fraudsters may use these to open bank accounts, obtain credit, or participate in social engineering campaigns.
- *Facial recognition spoofing*: Attackers can use AI to simulate or alter facial characteristics in photos or videos to impersonate legitimate users. Techniques such as face morphing allow the blending of two faces into a single image accepted by biometric systems, thereby enabling dual access [43].

4.3 Infrastructure security considerations

The incorporation of AI into digital identity systems also introduces broader infrastructure security challenges. These must be addressed to maintain the trustworthiness and resilience of the system. Some of these considerations are as follows:

- *Increased attack surfaces*: Each AI component, whether a model, API, or training pipeline; adds a new surface that can be targeted. In many cases, these components rely on third-party libraries or external datasets, which make them susceptible to supply chain attacks. AI-enabled identity services may also be exposed *via* APIs, which if poorly secured, can be exploited to conduct inference or denial-of-service attacks.
- *Complexity in system auditing and testing*: AI systems are inherently non-deterministic and opaque, making it difficult to audit or formally verify their behavior. This poses significant hurdles for identity system auditors and compliance officers. Traditional security testing methodologies are often inadequate for capturing the nuanced behaviors of learning systems. New techniques such as adversarial testing, fairness auditing, and explainability frameworks are required.
- *Deployment and lifecycle risks*: The AI lifecycle, spanning data collection, model training, deployment, and updating; introduces multiple points of failure. For example, identity systems that rely on continuous learning may incorporate malicious inputs post-deployment, resulting in concept drift or degraded accuracy over time. Ensuring the integrity of models throughout their lifecycle requires robust monitoring and incident response strategies.

- *Privacy-preserving infrastructure requirements*: Many AI-based identity systems process biometric and behavioral data. Infrastructure must therefore support privacy-preserving mechanisms such as differential privacy, homomorphic encryption, and federated learning. Without such safeguards, there is a risk of violating data protection laws and exposing users to privacy harm.

4.4 Strategies for enhancing AI security in identity systems

To mitigate the emerging threats associated with AI integration in identity systems, several best practices and technological strategies should be considered:

- *Model robustness testing*: Regularly test models against adversarial examples and simulate real-world spoofing attacks to assess resilience.
- *Secure model training pipelines*: Use trusted datasets, apply data provenance controls, and isolate training environments.
- *Zero trust architecture*: Extend the zero trust model to encompass AI components, ensuring strict identity verification for inter-process and inter-service communication.
- *Transparency and explainability tools*: Incorporate model explainability techniques to build trust and allow for human oversight of automated decisions.
- *Incident monitoring and response*: Develop AI-specific threat detection and response capabilities to handle model drift, poisoning attempts, or anomalous output patterns.

5. Privacy and risk management considerations

This section explores key privacy risks, regulatory responses, and risk management strategies relevant to AI-enhanced identity systems.

5.1 Privacy risks of AI inference and profiling

One of the most profound privacy concerns arises from AI's ability to infer sensitive attributes from seemingly benign identity data. Even when users share minimal personal information, AI models can extract detailed profiles by analyzing behavior, transaction patterns, device fingerprints, voice data, facial imagery, or network metadata. These inferences may include:

- Demographic details (e.g., age, gender, and ethnicity)
- Psychological traits (e.g., mood, personality, and political orientation)
- Health indicators (e.g., stress levels and genetic predisposition)
- Social or financial status

Such profiling, often done without explicit consent or knowledge of the user, challenges existing notions of informational self-determination and can lead to discrimination, exclusion from services, or even surveillance creep [44]. Furthermore, AI systems trained on historical identity data may embed societal biases, reinforcing discrimination against marginalized groups [45]. These biases are often opaque due to the black-box nature of many AI models, which complicates accountability and contestability for impacted individuals.

5.2 Trade-offs between personalization and privacy

Personalized services powered by identity-linked AI can improve user experience by adapting interfaces, access controls, or authentication steps based on user context. However, this personalization requires extensive data collection and real-time monitoring, which poses ethical dilemmas around autonomy and privacy. The core trade-off lies between:

- user convenience and service quality and
- user privacy and control over data processing.

In identity systems, this may manifest in adaptive authentication workflows where behavioral biometrics (e.g., keystroke dynamics) are analyzed continuously [26]. While effective for security, it may feel intrusive to users, unaware that their behaviors are being monitored and analyzed. Striking the right balance necessitates transparent communication, opt-in mechanisms, and proportional data usage aligned with privacy principles.

5.3 Privacy by design and emerging AI regulations

To address these concerns, regulatory frameworks are evolving to govern the use of AI in identity and data processing contexts. Notably:

- GDPR [28] mandates principles such as purpose limitation, data minimization, and explicit consent for profiling and automated decision-making. Articles 22 and 35 are relevant, requiring human oversight and Data Protection Impact Assessments (DPIAs) for high-risk AI applications.
- EU AI Act [46] introduces a risk-based regulatory structure classifying AI systems into unacceptable, high-risk, limited-risk, and minimal-risk categories. Biometric identification, remote emotion recognition, and AI used in law enforcement are deemed high-risk, demanding rigorous documentation, transparency, and human oversight mechanisms.
- NIST AI Risk Management Framework (AI RMF) [47] promotes a voluntary yet robust model for developing trustworthy AI. It emphasizes principles such as fairness, accountability, and resilience, and includes identity-specific risks like inferential harm, model misuse, and adversarial robustness.

In combination, these frameworks encourage embedding Privacy by Design (PbD) [48, 49] into the development lifecycle of AI-driven identity systems. This means adopting techniques such as:

- differential privacy to protect training data,
- federated learning to avoid central data aggregation,
- homomorphic encryption for secure data processing, and
- model explainability tools to meet auditability requirements.

5.4 Risk management strategies for identity providers and data controllers

As AI reshapes identity ecosystems, both IdPs and data controllers must evolve their risk management strategies. Below are essential components of a proactive privacy and security posture:

- Risk identification and categorization: Conduct AI-specific DPIAs to identify risks arising from inferential capabilities, automation, or bias. This involves mapping data flows to ensure clarity on what identity attributes are collected, inferred, retained, or shared.
- Governance and accountability: Establish clear governance structures assigning responsibility for AI model performance, fairness, and security. This requires maintaining audit trails and documenting AI decision logic, especially when identity verification impacts service eligibility.
- User-centric risk controls: Enable user access to meaningful explanations and the ability to opt out of AI-based profiling. This would need to provide consent dashboards that reflect the real-time status of identity data usage.
- Continuous monitoring and incident response: Monitor AI systems for drift, adversarial attacks, or emerging bias. It is also important to prepare incident response protocols tailored to privacy violations stemming from AI behavior or misclassification.
- Training and organizational culture: Train AI engineers and risk officers on privacy laws, fairness metrics, and responsible innovation. It is also important to foster a culture of privacy-respectful design and ethical AI deployment within identity service organizations.

6. Toward a security-centric framework for AI in identity systems

This section outlines a security-centric framework for the design, deployment, and governance of AI-augmented identity systems, thereby ensuring trust, resilience, and accountability.

6.1 Security-by-design for AI-driven identity systems

With the recent increase in the deployment of AI-driven identity systems, it has become imperative to rethink their security paradigms and architectural foundations. Traditional identity architectures were not designed to account for the dynamic, data-driven behavior of AI components. As such, existing security models must evolve

to accommodate the unique risks, capabilities, and uncertainties that AI introduces. The security-by-design (SbD) model has emerged as the security model that could help ensure trust, resilience, and accountability within the online identity ecosystem. The SbD principle requires embedding security considerations into every stage of a system's development lifecycle. When adapted to AI-integrated identity systems, SbD involves:

- Secure data collection and labeling: Ensuring that data sources are verified, data quality is maintained, and annotation processes do not introduce bias or vulnerability.
- Adversarial robustness: Developing models resilient to evasion, poisoning, and inference attacks using adversarial training and robust optimization techniques.
- Privacy-preserving computation: Incorporating techniques like federated learning, homomorphic encryption, or differential privacy to minimize the exposure of sensitive identity data.
- Model explainability tools: Integrating XAI mechanisms to detect anomalous outputs and ensure decisions can be interpreted and challenged by human operators.

Importantly, SbD for AI systems must go beyond static system hardening and include adaptive risk monitoring, as AI behavior may evolve over time due to online learning or shifts in input distributions.

6.2 Governance mechanisms for AI accountability and auditability

Trustworthy identity systems require more than just secure code, they demand transparent and auditable governance structures. Several governance mechanisms are key:

- Model documentation and datasheets: Standardized formats like Model Cards [50] and Data Sheets for Datasets [51] can help stakeholders understand the intended use, limitations, and risks of AI components used in identity systems.
- Audit trails and logging: All AI-driven decisions, especially in high-stakes identity verification processes, should be logged along with confidence scores, input parameters, and system context to allow for retrospective analysis.
- Impact assessments and red-teaming: Before deployment, models should undergo AI risk assessments (e.g., under the EU AI Act or NIST AI RMF) to evaluate potential harms, biases, or misuse scenarios. Red-teaming exercises can simulate adversarial attacks or unintended use cases.
- HITL oversight: High-risk decisions (e.g., blocking access to critical services) should incorporate human review to mitigate the effects of false positives or opaque outcomes.

6.3 A layered model for building resilient identity ecosystems

A layered security model can provide a blueprint for resilience by segmenting responsibilities and control points across multiple tiers of the AI-enabled identity

systems. At the foundation of this model lies the principle of SbD, which involves embedding protective mechanisms as described in subsection 6.1 into the architecture and workflows of the systems from the outset, rather than retrofitting them post-deployment. The proposed model for AI-augmented identity systems includes several layers as shown in **Figure 1**.

The data layer includes techniques such as secure data ingestion, data minimization, anonymization, adversarial robust testing, privacy-preserving ML, and logging controls, which are essential at the early stages of system development. These approaches ensure that the model's training process does not inadvertently expose sensitive identity data or become vulnerable to data poisoning or model inversion attacks.

Next, is the model layer that must be governed by practices that enforce algorithmic accountability. This includes regular auditing of models for fairness, bias, and security vulnerabilities, and the use of XAI techniques to ensure transparency in decision-making [52]. For example, when AI is used to assess biometric inputs or behavioral patterns, the rationale behind accept/reject decisions must be comprehensible to both system operators and end users. This is particularly vital in contexts where AI decisions impact access to services, rights, or entitlements. Incorporating mechanisms for HITL supervision can mitigate risks associated with opaque or erroneous automated decisions.

At the decision layer, AI models produce actionable outputs that drive identity-related processes such as access control, anomaly detection, and identity federation. Security-centric design at this level involves enforcing strong access policies, logging

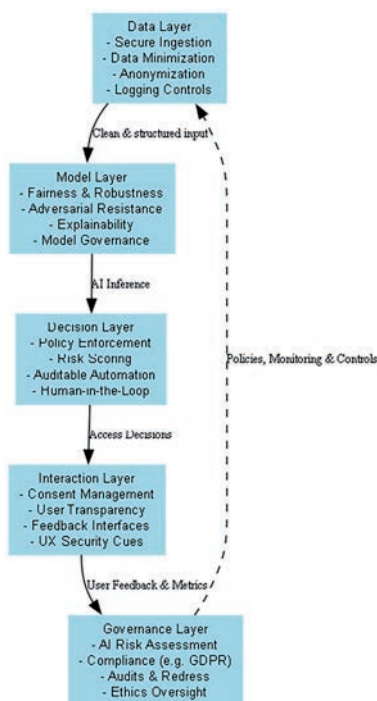


Figure 1.
Model for AI-augmented identity systems.

decision trails, and embedding mechanisms to flag anomalous or high-risk actions for human review. Systems must also implement thresholds and safeguards that prevent over-reliance on AI, especially in critical sectors such as healthcare, finance, and public services [53]. Decision outputs must be auditable and reversible, enabling accountability in case of wrongful identifications or breaches.

The interaction layer represents the point of contact between identity systems and their users. Security-centric design in this layer prioritizes user agency, consent, and transparency. Interfaces must clearly communicate how AI processes identity data, what decisions are being made, and what data is being collected. Usable privacy and security cues, such as dashboards that show data access patterns or explain consent status, can build user trust and encourage responsible data-sharing behavior [54]. Mechanisms that allow users to contest or appeal AI-driven identity decisions are also necessary for maintaining fairness and legitimacy.

Lastly, the governance layer provides overarching accountability, oversight, and strategic alignment with legal and ethical norms. This includes developing organizational AI policies that comply with regulatory frameworks such as the GDPR, the EU AI Act, and the AI RMF. It also involves establishing AI ethics boards, conducting algorithmic impact assessments, and maintaining documentation of model training processes, datasets, and validation results. Risk governance must be dynamic, such that it can adapt to evolving threat landscapes as well as incorporate continuous feedback from audits, user reports, and incident responses.

It is important to note that the framework emphasizes resilience and adaptability across all layers. This means that AI-augmented identity systems must be capable of detecting, responding to, and recovering from security incidents swiftly and effectively. Strategies such as layered redundancy, fail-safe protocols, and cross-domain incident coordination enhance the robustness of these systems. Furthermore, the integration of AI into identity infrastructures must be viewed as a long-term transformation, necessitating collaboration across disciplines; from cybersecurity and legal experts to ethicists and social scientists.

6.4 Strategic recommendations

The interplay between AI, security, and identity demands holistic strategies that span technical, organizational, and policy domains. The following recommendations are targeted at various stakeholders.

For governments and regulators:

- Develop AI-specific identity regulations that align with broader digital rights frameworks (e.g., GDPR and EU AI Act).
- Promote standards for biometric verification, identity inference, and AI-driven authentication tools.
- Fund independent audits and public testbeds for identity AI models to promote transparency.

For enterprises and identity providers:

- Adopt layered defenses across AI pipelines with routine adversarial testing and robustness validation.

- Prioritize vendor transparency when integrating third-party AI identity services, demand explainability and model documentation.
- Train internal teams in AI security, risk management, and governance principles.

For researchers and developers:

- Advance privacy-preserving identity research, including synthetic identity detection and federated identity modeling.
- Contribute to open benchmarks and datasets that facilitate fair and secure identity AI development.
- Explore ethical implications of persistent digital identities enhanced by continuous AI profiling.

7. Case studies and applications

To better understand the real-world implications of integrating AI into digital identity systems, this section presents a comparative analysis of three significant use cases across government, financial, and healthcare domains. Each example highlights the transformative potential of AI, while also revealing specific challenges related to security, privacy, and trust.

7.1 Example 1: Biometric authentication in National Identification Programs

Many countries have adopted national identification (ID) programs as a way to enhance secure access to public services and improve administrative efficiency [55, 56]. For instance, India's Aadhaar system, which is arguably the largest biometric ID program globally, uses AI algorithms to manage the distribution of government benefits [57]. The system operates by using ML to divide each household into different categories according to their economic situation, where each category is entitled to a specific amount of subsidized grains and commodities.

AI also enhances operational scalability by automating verification tasks and detecting anomalies in biometric data patterns, such as matching inconsistencies or spoofing attempts [58]. However, the opaque nature of some biometric matching algorithms raises concerns about accuracy, bias, and explainability, particularly for marginalized populations with higher biometric failure rates [59]. These programs illustrate the need for accountable AI deployment with built-in mechanisms for contestability and error correction.

7.2 Example 2: AI for fraud detection in financial services

In the financial sector, AI has become integral to identity verification and fraud prevention strategies. Banks and fintech platforms routinely use machine learning models to analyze behavioral data, device fingerprints, transaction history, and social graph signals to flag suspicious activities [60]. For instance, AI-based identity verification solutions can utilize computer vision and natural language processing to validate identity documents and detect forgeries in real time [61].

These models significantly reduce manual workload and fraud losses by identifying synthetic identities, account takeovers, and money laundering patterns with high precision. However, the black-box nature of many fraud detection systems can pose regulatory and ethical challenges, especially when legitimate customers are wrongly flagged [62]. As a result, organizations are increasingly being asked to demonstrate the fairness, proportionality, and auditability of their AI models to comply with frameworks like GDPR and the NIST AI RMF. Also, the use of AI for continuous behavioral monitoring raises privacy concerns, particularly around the secondary use of identity attributes for marketing or profiling [63]. This case emphasizes the trade-off between AI-driven vigilance and the need for user-centric, transparent risk management.

7.3 Example 3: Digital Onboarding in healthcare

In the healthcare domain, AI plays a critical role in digital onboarding processes, especially in remote or telemedicine contexts [64]. Identity verification in patient portals, insurance systems, and e-prescription platforms often leverages AI-based facial recognition, document analysis, and behavioral verification. This has become especially relevant in the post-COVID landscape, where the demand for contactless patient interaction surged.

Despite the benefits, such applications raise privacy and security concerns due to the sensitivity of health data. Inference risks, where AI systems deduce medical conditions or vulnerabilities from identity information, underscore the importance of integrating privacy-preserving techniques into AI solutions in healthcare [65]. Moreover, the sector's reliance on third-party identity verification providers introduces additional dependencies and regulatory exposure, especially under GDPR.

7.4 Comparative insights

Across all three sectors, AI augments identity systems by enabling real-time decision-making, scalability, and adaptive security mechanisms. Yet, each application domain reveals unique tensions between AI utility and responsible identity management. National ID programs must grapple with inclusivity and algorithmic fairness; financial institutions must balance fraud detection with user rights; and healthcare systems must prioritize data minimization and consent.

These cases collectively demonstrate the urgent need for a context-aware, security-centric governance model that aligns AI deployment with societal values, legal obligations, and user expectations. As identity becomes increasingly digital and intelligent, the lessons drawn from these applications will shape the future trajectory of secure, trustworthy identity ecosystems.

8. Future directions and research opportunities

Several emerging developments promise to redefine how identity is created, managed, and secured. One notable frontier is the integration of quantum-safe cryptographic techniques into identity systems. With the anticipated advent of practical quantum computing, current cryptographic primitives underpinning digital identity risk becoming obsolete. Research into quantum-resistant algorithms and their incorporation with AI-powered identity verification mechanisms is critical to future-proof identity ecosystems. This transition will require close coordination between

cryptographers, AI experts, and system architects to maintain security without compromising performance or usability.

Another domain demanding urgent attention is the rise of generative AI risks. Generative models capable of producing highly realistic synthetic identities, documents, and biometric traits (e.g., deepfakes) threaten to undermine trust in digital identity frameworks. Addressing these risks will involve developing advanced detection algorithms, robust provenance tracing, and embedding forensic AI capabilities into identity verification workflows. Moreover, governance structures must evolve to regulate generative AI use responsibly while enabling innovation.

The concept of decentralized AI agents also opens promising avenues for identity management. Autonomous agents running on decentralized networks could offer more user-centric control over identity data, enabling privacy-preserving credential issuance, selective disclosure, and peer-to-peer trust models. These paradigms could synergize with SSI initiatives to shift power back to individuals while maintaining verifiability and resilience. However, this requires interdisciplinary research spanning AI, blockchain, cryptography, and human-computer interaction to overcome scalability, governance, and usability challenges.

To fully harness these advancements, the field must embrace interdisciplinary collaboration involving computer scientists, legal scholars, social scientists, ethicists, and policymakers. The complex interplay between technical innovation, regulatory frameworks, and social acceptance necessitates a holistic approach to research and development. Also, there is a pressing need for longitudinal studies tracking the deployment and societal impact of AI-enhanced identity systems over time. Such research can provide empirical evidence on system effectiveness, unintended consequences, and evolving user perceptions, informing adaptive policy-making.

Moreover, establishing policy benchmarking frameworks across jurisdictions will enable stakeholders to compare regulatory approaches, identify best practices, and harmonize standards for AI governance in identity systems. This global coordination is vital to ensuring interoperability, legal certainty, and protection of fundamental rights in an increasingly connected digital world. In all, the future of AI in online identity is rich with potential but fraught with challenges that demand proactive, collaborative, and ethical approaches to research, development, and governance.

9. Conclusion

This chapter has explored the multifaceted role of AI in shaping the evolution of online identity systems. Beginning with foundational architectures and progressing through trust dynamics, security implications, privacy considerations, and governance frameworks, the analysis highlighted both the transformative potential and inherent risks of integrating AI into identity management. Also, AI's ability to enhance scalability, accuracy, and adaptability in identity verification and fraud detection offers unprecedented opportunities to build more secure and user-friendly digital ecosystems. However, these benefits come with challenges, particularly around the opacity of AI algorithms, emerging threats such as adversarial attacks and synthetic identity fraud, and complex trade-offs between personalization and privacy.

Trust remains the cornerstone of any identity system, and AI introduces new dimensions to this trust relationship among users, identity providers, and automated

systems. Achieving explainability, accountability, and user-centric control are essential to sustaining confidence in AI-driven identity technologies. The ethical considerations discussed throughout this chapter emphasize the need for SbD principles, robust governance mechanisms, and interdisciplinary collaboration to guide the responsible development and deployment of AI in identity systems. As emerging technologies such as quantum-safe cryptography and decentralized AI agents advance, ongoing research and policy innovation will be crucial to navigating the evolving landscape. Thus, fostering a secure, transparent, and inclusive AI-augmented identity ecosystem will require a shared commitment from technologists, regulators, and users alike. Through balancing innovation with ethical stewardship, the future of online identity can be one that empowers individuals while safeguarding fundamental rights and societal trust.

Author details

Livinus Obiora Nweke^{1,2}

1 Noroff University College, Kristiansand, Norway

2 University of South-Eastern Norway, Borre, Norway

*Address all correspondence to: livinus.nweke@noroff.no

IntechOpen

© 2025 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Zhu X, Badr Y. Identity management systems for the internet of things: A survey towards blockchain solutions. *Sensors*. 2018;**18**(12):4215
- [2] Arner DW, Zetzsche DA, Buckley RP, Barberis JN. The identity challenge in finance: From analogue identity to digitized identification to digital kyc utilities. *European Business Organization Law Review*. 2019;**20**(1):55-80
- [3] El Haddouti S, El Kettani MDE-C. Analysis of identity management systems using blockchain technology. In: 2019 International Conference on Advanced Communication Technologies and Networking (CommNet). Rabat, Morocco: IEEE; 2019. pp. 1-7
- [4] Nweke L, O. National Identification Systems As Enablers of Online Identity. London, UK: IntechOpen; 2023
- [5] Keerthana V, Sneha M, Lokeshwaran A, Pranesh K. Adaptive fraud detection in digital banking using deep learning: A hybrid cnn-rnn approach. In: 2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES). Chennai, India: IEEE; 2024. pp. 1-7
- [6] Tamraparani V. Leveraging ai for fraud detection in identity and access management: A focus on large-scale customer data. *SSRN Electronic Journal*. 2025;**31**(4):719-727
- [7] Vegas J, Llamas C. Opportunities and challenges of artificial intelligence applied to identity and access management in industrial environments. *Future Internet*. 2024;**16**(12):469
- [8] Habbal A, Ali MK, Abuzaraida MA. Artificial intelligence trust, risk and security management (ai trism): Frameworks, applications, challenges and future research directions. *Expert Systems with Applications*. 2024;**240**:122442
- [9] Bodó B. Mediated trust: A theoretical framework to address the trustworthiness of technological trust mediators. *New Media and Society*. 2020;**23**(9):2668-2690
- [10] Balan A, Tan AG, Kourtiti K, Nijkamp P. Data-driven intelligent platforms—design of self-sovereign data trust systems. *Land*. 2023;**12**(6):1224
- [11] Goel A, Rahulamathavan Y. A comparative survey of centralised and decentralised identity management systems: Analysing scalability, security, and feasibility. *Future Internet*. 2024;**17**(1):1
- [12] Carretero J, Izquierdo-Moreno G, Vasile-Cabezas M, Garcia-Blas J. Federated identity architecture of the european eid system. *IEEE Access*. 2018;**6**:75302-75326
- [13] Cucko S, Turkanovic M. Decentralized and self-sovereign identity: Systematic mapping study. *IEEE Access*. 2021;**9**:139009-139027
- [14] Vo TH, Fuhrmann W, Fischer-Hellmann K-P, Furnell S. Identity-as-a-service: An adaptive security infrastructure and privacy-preserving user identity for the cloud environment. *Future Internet*. 2019;**11**(5):116
- [15] AlRousan M, Intrigila B. Multi-factor authentication for e-government services using a smartphone application and biometric identity verification. *Journal of Computer Science*. 2020;**16**(2):217-224

- [16] Tiwari S, Raja R, Wadawadagi RS, Naithani K, Raja H, Ingle D. Emerging Biometric Modalities and Integration Challenges, Chapter Online Identity-an Essential Guide. London, UK: IntechOpen; 2023
- [17] Xiaochen H, Zhang X, Lovrich NP. Forecasting identity theft victims: Analyzing characteristics and preventive actions through machine learning approaches. *Victims and Offenders*. 2020;**16**(4):465-494
- [18] Felouat H, Nguyen HH, Le T-N, Yamagishi J, Echizen I. Ekyc-df: A large-scale deepfake dataset for developing and evaluating ekyc systems. *IEEE Access*. 2024;**12**:30876-30892
- [19] Martín AG, Beltrán M, Fernández-Isabel A, Martín I, de Diego. An approach to detect user behaviour anomalies within identity federations. *Computers and Security*. 2021;**108**:102356
- [20] Buolamwini J, Gebru T. Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, PMLR. 2018;**81**:77-91
- [21] Huang K, Narajala VS, Yeoh J, Ross J, Raskar R et al. A Novel Zero-Trust Identity Framework for Agentic AI: Decentralized Authentication and Fine-Grained Access Control. 2025
- [22] Vadim S, Firdaus M, Rhee K-H. Privacy-preserving decentralized biometric identity verification in car-sharing system. *Journal of Multimedia Information System*. 2024;**11**(1):17-34
- [23] Vakhshiteh F, Nickabadi A, Ramachandra R. Adversarial attacks against face recognition: A comprehensive study. *IEEE Access*. 2021;**9**:92735-92756
- [24] Ferrara E. Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *The Sciences*. 2023;**6**(1):3
- [25] Nweke LO, Yayilgan SY. Opportunities and Challenges of Using Artificial Intelligence in Securing Cyber-Physical Systems. Switzerland: Springer Nature; 2024. pp. 131-164
- [26] Krishnamoorthy S, Rueda L, Saad S, Elmiligi H. Identification of user behavioral biometrics for authentication using keystroke dynamics and machine learning. In: *Proceedings of the 2018 2nd International Conference on Biometric Engineering and Applications, ICBEA '18*. Amsterdam, Netherlands: ACM; 2018. pp. 50-57
- [27] Kaur D, Uslu S, Rittichier KJ, Durresi A. Trustworthy artificial intelligence: A review. *ACM Computing Surveys*. 2022;**55**(2):1-38
- [28] European Union. Regulation (eu) 2016/679 of the european parliament and of the council. *Official Journal of the European Union*. 2016:6-8
- [29] Wachter S, Mittelstadt B, Russell C. Counterfactual explanations without opening the black box: Automated decisions and the gdpr. *Harvard Journal of Law and Technology*. 2018;**31**(2):849-853
- [30] Krishnapriya KS, Albiero V, Vangara K, King MC, Bowyer KW. Issues related to face recognition accuracy varying based on race and skin tone. *IEEE Transactions on Technology and Society*. 2020;**1**(1):8-20
- [31] Cavazos JG, Jonathon Phillips P, Castillo CD, O'Toole AJ. Accuracy comparison across face recognition algorithms: Where are we on measuring race bias? *IEEE Transactions on*

Biometrics, Behavior, and Identity Science. 2021;3(1):101-111

[32] Grother P, Ngan M, Hanaoka K. Face Recognition Vendor Test Part 3: Demographic Effects. 2019

[33] Lundberg S, Lee Su-In. A Unified Approach to Interpreting Model Predictions. 2017

[34] Ribeiro M, T, Singh S, Guestrin C. Why should i trust you?: Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. San Francisco, California, USA: ACM; 2016. pp. 1135-1144

[35] Nass C, Moon Y. Machines and mindlessness: Social responses to computers. Journal of Social Issues. 2000;56(1):81-103

[36] Christiaens T. Trust and power in airbnb's digital rating and reputation system. Ethics and Information Technology. 2025;27(2):1-13

[37] Wiyatno RR, Xu A, Dia O, de Berker A. Adversarial Examples in Modern Machine Learning: A Review. 2019

[38] Wang Z, Ma J, Wang X, Jiahui H, Qin Z, Ren K. Threats to training: A survey of poisoning attacks and defenses on machine learning systems. ACM Computing Surveys. 2022;55(7):1-36

[39] Tolpegin V, Truex S, Gursoy ME, Liu L. Data Poisoning Attacks against Federated Learning Systems. Switzerland: Springer International Publishing; 2020. pp. 480-501

[40] Zhang Y, Jia R, Pei H, Wang W, Li B, Song D. The secret revealer: Generative model-inversion attacks against deep

neural networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE; 2020. pp. 250-258

[41] Korshunov P, Marcel S. Deepfakes: A New Threat to Face Recognition? Assessment and Detection. 2018

[42] Ghiurău D, Popescu DE. Distinguishing reality from ai: Approaches for detecting synthetic content. Compute. 2024;14(1):1

[43] Venkatesh S, Ramachandra R, Raja K, Busch C. Face morphing attack generation and detection: A comprehensive survey. IEEE Transactions on Technology and Society. 2021;2(3):128-145

[44] Veale M, Edwards L. Clarity, surprises, and further questions in the article 29 working party draft guidance on automated decision-making and profiling. Computer Law and Security Review. 2018;34(2):398-404

[45] Ntoutsis E, Fafalios P, Gadiraju U, Iosifidis V, Nejdil W, Vidal M-E, et al. Bias in data-driven artificial intelligence systems—An introductory survey. WIREs Data Mining and Knowledge Discovery. 2020;10(3):1-4

[46] European Union. Regulation (eu) 2024/1689 of the european parliament and of the council. Official Journal of the European Union. 2024:8-15

[47] NIST. Artificial Intelligence Risk Management framework (Ai Rmf 1.0). Technical report. Gaithersburg, MD, USA: National Institute of Standards and Technology; 2023

[48] Cavoukian A. Privacy by design: The 7 foundational principles. Information and Privacy Commissioner of Ontario. 2009;5:2009

- [49] Nweke LO. Using the Seven Principles of Privacy by Design (pbd) to Explain the Benefits of pbd in the Development of National Identification Systems. 2023
- [50] Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al. Model cards for model reporting. In: Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19. Atlanta, GA, USA: ACM; 2019. pp. 220-229
- [51] Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Hal Daumé III, et al. Datasheets for datasets. *Communications of the ACM*. 2021;**64**(12):86-92
- [52] Doshi-Velez F, Kim B. Towards a Rigorous Science of Interpretable Machine Learning. 2017
- [53] Morley J, Machado CCV, Burr C, Cows J, Joshi I, Taddeo M, et al. The ethics of ai in health care: A mapping review. *Social Science and Medicine*. 2020;**260**:113172
- [54] Utz C, Degeling M, Fahl S, Schaub F, Holz T. (un)informed consent: Studying gdpr consent notices in the field. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS '19. London, UK: ACM; 2019. pp. 973-990
- [55] Nweke L, O, Abomhara M, Yayilgan SY, Comparin D, Heurtier O, Bunney C. A LINDDUN-based privacy threat modelling for national identification systems. In: 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON). Abuja, Nigeria: IEEE; 2022
- [56] Abomhara M, Nweke LO, Yayilgan SY, Comparin D, Teyras K, de Labriolle S. Enhancing privacy protections in national identification systems: An examination of stakeholders' knowledge, attitudes, and practices of privacy by design. *International Journal of Information Security*. 2024;**23**(6):3665-3689
- [57] Srinivasalu S, Selladurai M, Sharma S, Perumal G, Mathaiyan M, Jo AA, et al. Social Assessment and Cultural Resistance: The Public Distribution System in Tamil Nadu, India. Switzerland: Springer Nature; 2025. pp. 169-185
- [58] Mir U, Kar AK, Gupta MP. Ai-enabled digital identity – Inputs for stakeholders and policymakers. *Journal of Science and Technology Policy Management*. 2021;**13**(3):514-541
- [59] Drozdowski P, Rathgeb C, Dantcheva A, Damer N, Busch C. Demographic bias in biometrics: A survey on an emerging challenge. *IEEE Transactions on Technology and Society*. 2020;**1**(2):89-103
- [60] Kumar S, Ahmed R, Bharany S, Shuaib M, Ahmad T, Eldin ET, et al. Exploitation of machine learning algorithms for detecting financial crimes based on customers' behavior. *Sustainability*. 2022;**14**(21):13875
- [61] Makandar A, Kaman S, Javeriya SB. Combating digital forgeries: Advanced ai techniques for detecting forgeries of diverse data. *AI, Computer Science and Robotics Technology*. 2024;**3**:2-4
- [62] Nesvijejskaia A, Ouillade S, Guilmin P, Zucker J-D. The accuracy versus interpretability trade-off in fraud detection model. *Data and Policy*. 2021;**3**:e12-3-e12-6
- [63] Alhitmi H, K, Mardiah A, Al-Sulaiti KI, Abbas J. Data security

and privacy concerns of ai-driven marketing in the context of economics and business field: An exploration into possible solutions. *Cogent Business and Management*. 2024;**11**(1):4-9

[64] Carrie JC, Winter S, Steiner D, Wilcox L, Terry M. Hello ai: Uncovering the onboarding needs of medical practitioners for human-ai collaborative decision-making. *Proceedings of the ACM on Human-Computer Interaction*. 2019;**3**(CSCW):1-24

[65] Khalid N, Qayyum A, Bilal M, Al-Fuqaha A, Qadir J. Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine*. 2023;**158**:106848



Edited by Livinus Obiora Nweke

In a world increasingly defined by artificial intelligence (AI), digital identity, and algorithmic trust, our sense of self is no longer confined to the physical realm; it extends, adapts, and evolves across data-driven platforms. This book presents an interdisciplinary examination of how intelligent systems are transforming the foundations of identity, privacy, and autonomy in the digital age. Bridging perspectives from psychology, education, information security, and governance, this volume examines the emerging intersections between human behavior, ethical design, and advanced technologies, including EdgeAI, federated learning, and self-sovereign identity. Through accessible analysis and forward-looking discussion, it reveals how online identities are constructed, authenticated, and safeguarded in an era of rapid technological transformation. The book's strength lies in its ability to connect technical architecture with human experience, showing how trust, transparency, and digital resilience must be intentionally engineered into identity ecosystems. Readers will gain an integrated understanding of the psychological, social, and regulatory dimensions that underpin secure and ethical digital participation. Comprehensive yet approachable, this volume serves as a valuable reference for researchers, educators, technologists, and policymakers seeking to navigate the challenges of identity management in an AI-driven world. By combining conceptual clarity with practical insight, it invites reflection on a crucial question for the future: how can we build identity systems that recognize not only who we are, but what we value?

Published in London, UK
© 2026 IntechOpen
© vsijan / nightcafe.studio

IntechOpen

