



FACULTY OF LAW

LUND UNIVERSITY

Nomin Soyol-Erdene

AI and the Inner Sphere: Privacy and Proportionality under International Human Rights Law

JAMM07 Master Thesis

International Human Rights Law

30 higher education credits

Supervisor: Matthew Scott

Term of graduation: Spring 2026

Table of contents

Summary	4
Preface.....	5
Abbreviations	7
1 Introduction.....	8
1.1 Research Objectives and Research Questions	9
1.2 Methodology.....	10
1.3 Scope and Limitations	11
1.4 Outline	11
2 Privacy, Proportionality, and the “Inner Sphere” Under International Human Rights Law	13
2.1 The Right to Privacy.....	13
2.1.1 The Legality Requirement.....	15
2.1.2 Heightened Standards for Technology	16
2.1.3 The Proportionality Test	17
2.2 Privacy Beyond Informational Secrecy.....	18
2.3 The “Inner Sphere”	20
2.4 Proportionality Under International Human Rights Law	22
2.5 State Obligations, Private Actors, and the UN Guiding Principles	23
2.5.1 The Doctrinal Basis of Positive Obligations.....	24
2.5.2 The UN Guiding Principles.....	24
2.5.2.1 The State Duty to Protect	25
2.5.2.2 The Corporate Responsibility to Respect	25
2.5.2.3 Access to Remedy	26
3 AI, Privacy, and the Structural Erosion of the Private Sphere	28
3.1 From Informational Privacy to Behavioral Inference	29
3.1.1 Digital Phenotyping	29
3.1.2 Predictive Analytics	30
3.1.3 Emotional Inference	31
3.1.4 Boundary Management	32
3.2 Chilling Effects and Behavioral Modification	32
3.3 Algorithmic Vulnerability	33
3.3.1 Children and Youth in the Algorithmic System.....	34
3.3.2 Adults in Crisis.....	35
4 State Responsibility in AI Governance.....	37

4.1	Negative Obligation: <i>The Dutch Systeem Risicoindicatie</i> Case	37
4.1.1	The Court’s Reasoning.....	39
4.2	Positive Obligation: The Bărbulescu Proportionality Test..	42
4.3	Synthesis: Structural Harms and the Proportionality Gap...	44
4.4	AI Governance and the Limits of Current Regulatory Frameworks	46
4.4.1	The Transition to Binding Human-Rights-Based Regulation	47
4.4.2	The “Salient Harm” Model	48
4.4.3	The Resilience of the Inner Sphere	49
4.4.4	Operationalizing Human Rights Due Diligence as a Tool for Structural Scrutiny.....	49
5	Conclusion	53
6	Bibliography	56

Summary

This thesis examines how contemporary artificial intelligence systems challenge the protection of the right to privacy under international human rights law, particularly Article 17 of the International Covenant on Civil and Political Rights and Article 8 of the European Convention on Human Rights. It argues that privacy should not be understood only as protection against unlawful access to personal information, but as a legal safeguard for the individual's inner sphere, including autonomy, mental integrity, identity formation, and autonomous self-development.

The central argument advanced in this thesis is that AI systems interfere with privacy in ways that go beyond traditional forms of data collection or disclosure. Through behavioural inference, predictive analytics, emotional profiling, digital phenotyping, personalised recommendation systems, and conversational AI, such systems increasingly shape the environments within which individuals think, decide, behave, and develop identity. These interferences are often not visible, immediate, or easily measurable. Instead, they operate through psychological pressure, anticipatory conformity, chilling effects, emotional dependency, and gradual behavioural modification.

Therefore, existing proportionality analysis under international privacy law remains under-inclusive when it fails to recognise cumulative, psychological, and structurally embedded harms as legally relevant interferences with privacy. Although the existing doctrine already contains important mechanisms, including legality, foreseeability, necessity, safeguards, fair balance, and positive obligations, its current application remains too closely attached to a model of salient harm. This model is better suited to visible, direct, and procedurally identifiable privacy violations than to the slow and diffuse harms produced by AI-mediated environments.

Preface

It has been a rich experience.

Looking back at myself two years ago, I can see how much I have changed, even in ways I did not fully notice while they were happening. I used to think I knew what people meant when they described something as an “eye-opening experience.” I thought I understood it as an idea. But in reality, I had not truly understood it until I lived through these two years. Now I realise that an eye-opening experience is not always one sudden moment of clarity. Sometimes, it is the slow accumulation of everything: the difficult days, the quiet mornings, the people who stay, the people who leave, the conversations that change something in you, and the small moments that gradually teach you how to see yourself, others, and life differently.

My time in Lund has given me many kinds of experiences. Some were beautiful, some were painful, and some were simply confusing while I was living through them. But together, they taught me something I could not have learned only from books. Life goes on. The sun comes out again. As long as we are still breathing, there is still a way forward. Even when things feel heavy, there are still small things that make life softer: a kind word, a familiar place, the comfort of friendship, and the simple beauty of being able to relate to another person.

Perhaps one of the most important things I have learned during these two years is to appreciate those little things. I am deeply grateful for the friendships and connections I made here, and for the people who made Lund feel not only like a place where I studied, but also like a place where I grew.

I would like to express my sincere gratitude to my supervisor, Matthew Scott, for his guidance, patience, trust, and encouragement throughout this process. His support meant a great deal to me, especially during moments when I felt uncertain.

Also, to the Faculty of Law, and to every faculty member who was part of this journey. I still remember the first time Jessica told us, “You will learn a lot, but most importantly, you will learn a lot from each other.” At the time, I understood those words, but now I know how true they were. I learned so much not only from the programme itself, but also from my fellow students and friends, through conversations, different perspectives, and the simple experience of growing beside one another. I would especially like to thank my dearest friends, Andria and Anastasiia, for their kindness, support, and for helping me in ways I will always remember. Thank you to everyone who made these two years meaningful.

I am also deeply grateful to my family. In particular, I want to thank my dear younger uncle, Bayan, for giving me the opportunity to pursue this path and for supporting me in ways that made these two years possible. I also send my love and gratitude to my grandmother and grandfather, who raised me and shaped the person I am today. Their love has always been part of the foundation I carry with me.

- Nomin

Abbreviations

AI	Artificial Intelligence
ECHR	European Convention on Human Rights
ECtHR	European Court of Human Rights
FRIAs	Fundamental Rights Impact Assessments
HRC	The Human Rights Committee
HRDD	Human Rights Due Diligence
ICCPR	International Covenant on Civil and Political Rights
IHRL	International Human Rights Law
OHCHR	The Office of the High Commissioner for Human Rights
<i>SyRI</i>	<i>NJCM c.s. and others v De Staat der Nederlanden</i>
UNGP	The UN Guiding Principles on Business and Human Rights

1 Introduction

The rapid transition from industrialism to informationalism has fundamentally altered the structural conditions of human interaction.¹ Artificial intelligence (AI) systems increasingly mediate communication, decision-making, emotional support, access to information, and processes of identity formation.² Unlike traditional information technologies, contemporary AI systems do not merely facilitate the collection, storage, or exchange of data. Through machine learning, large language models, predictive analytics, and recommender systems, they are increasingly capable of identifying patterns, generating inferences, and shaping the informational environments within which individuals think, choose, and act.³

This technological shift has enabled what may be described as an “inference economy”, where value lies not only in the raw data collected, but in the sensitive insights derived from it.⁴ AI systems may infer emotional states, latent vulnerabilities, behavioral tendencies, and future conduct from seemingly ordinary or innocuous digital traces.⁵ Through techniques such as digital phenotyping, predictive profiling, and emotional inference, individuals may be categorized, evaluated, and targeted for commercial, political, administrative, or security purposes, often without meaningful awareness or control.⁶

These developments challenge the traditional privacy model within international human rights law. Privacy under Article 17 International Covenant on Civil and Political Rights (ICCPR) and Article 8 European Convention on Human Rights (ECHR) has often been analyzed through interferences that are external, observable, and materially visible, such as surveillance, data collection, disclosure, or monitoring.⁷ However, AI-mediated interferences do not always take the form of a discrete or temporally proximate intrusion. They may instead operate through opacity, behavioral inference, environmental modulation, and the gradual shaping of the conditions under which autonomy is exercised.⁸

¹ Julie E Cohen, ‘Surveillance vs. Privacy: Effects and Implications’ in David Gray and Stephen E Henderson (eds), *The Cambridge Handbook of Surveillance Law* (Cambridge University Press 2017) 455.

² Sue Anne Teo, ‘How to think about freedom of thought (and opinion) in the age of AI’ (2024) 53 *Computer Law & Security Review* 105969. 2

³ Cohen, (n 1) p. 459.

⁴ Teo, (n 2) p.5.

⁵ *ibid.*, p. 2.

⁶ Lisa Cosgrove et al, ‘Digital Phenotyping and Digital Psychotropic Drugs: Mental Health Surveillance Tools That Threaten Human Rights’ (2020) 22(2) *Health and Human Rights Journal*, p. 34.

⁷ Sue Anne Teo, ‘How Artificial Intelligence Systems Challenge the Conceptual Foundations of the Human Rights Legal Framework’ (2022) 40(1) *Nordic Journal of Human Rights*, p. 216.

⁸ *ibid.*, p. 230.

In this context, the individual’s “inner sphere” becomes increasingly vulnerable.⁹ AI systems may affect not only what is known about a person, but also the conditions under which that person forms preferences, develops identity, manages boundaries, and makes decisions. Persistent algorithmic observation and evaluation may produce chilling effects, psychological pressure, and anticipatory conformity, as individuals adjust their behavior in response to the possibility of being profiled, ranked, or negatively categorized.¹⁰ The harm is therefore not only informational, but also cumulative, psychological, and structurally embedded.

This thesis argues that international human rights law must take these forms of harm seriously. The problem is not necessarily that existing privacy doctrine lacks the tools to address AI-mediated interferences. Rather, the difficulty lies in the way proportionality analysis often remains attached to a model of salient harm, prioritizing visible, immediate, and procedurally identifiable violations.¹¹ As a result, cumulative and psychologically mediated erosions of autonomy, behavioral freedom, and identity formation may remain insufficiently recognized within existing privacy analysis.¹²

1.1 Research Objectives and Research Questions

The primary objective of this thesis is to examine whether existing international human rights privacy doctrine, specifically under Article 17 ICCPR and Article 8 ECHR, is capable of addressing AI-mediated harms that operate through inference, behavioral modification, and structural modulation. While both provisions protect the individual against unjustified interferences with privacy, they do so through different doctrinal structures.¹³ Article 8 ECHR contains an explicit limitations clause requiring that interferences be in accordance with the law, pursue a legitimate aim, and be necessary in a democratic society.¹⁴ Article 17 ICCPR prohibits arbitrary or unlawful interferences, with arbitrariness understood to include requirements of reasonableness, necessity, and proportionality.¹⁵

This thesis examines whether these frameworks can adequately respond to the “proportionality gap” created by AI systems.¹⁶ This gap arises where legal analysis recognizes formal requirements, but fails to fully account for the

⁹ Teo, (n 2) p. 9.

¹⁰ Jonathan W Penney, ‘Understanding Chilling Effects’ (2022) 106(3) *Minnesota Law Review* pp. 1459, 1491.

¹¹ Teo, (n 7) p. 218.

¹² *ibid.*, p. 234.

¹³ Sarah Joseph and Melissa Castan, *The International Covenant on Civil and Political Rights: Cases, Materials, and Commentary* (3rd edn, Oxford University Press 2013) p. 538.

¹⁴ *ibid.*

¹⁵ UN Human Rights Council, *The Right to Privacy in the Digital Age: Report of the United Nations High Commissioner for Human Rights* UN Doc A/HRC/48/31 (13 September 2021) para 8.

¹⁶ Teo, (n 7) p. 216.

substantive effects of AI systems on autonomy, mental self-determination, behavioral freedom, and the individual's inner sphere.¹⁷

The central research question is:

To what extent can proportionality analysis under Article 17 ICCPR and Article 8 ECHR address AI-mediated harms to privacy, autonomy, and the individual's inner sphere?

This central question is supported by the following sub-questions:

1. How is the relationship between privacy, personal autonomy, identity formation, and the inner sphere articulated within international human rights law?
2. In what ways do AI capabilities such as behavioral inference, emotional profiling, digital phenotyping, and environmental modulation constitute interferences with privacy?
3. To what extent does existing proportionality doctrine recognise cumulative and psychological harms?
4. What positive obligations do States have to protect individuals against AI-mediated privacy interferences by private actors?

1.2 Methodology

This thesis employs a doctrinal legal methodology and critical doctrinal analysis, to examine the nature of privacy interferences in the algorithmic age. The research systematises and interprets primary legal sources, including Article 17 ICCPR and Article 8 ECHR, together with interpretive materials such as HRC General Comment No 16. The analysis further draws upon ECtHR case law, UN and OHCHR reports, and Council of Europe materials in order to identify the standards of legality, foreseeability, necessity, proportionality, safeguards, and positive obligations applicable to digital surveillance, behavioral monitoring, and AI-mediated interferences with private life.

The doctrinal inquiry is supported by engagement with the UN Guiding Principles on Business and Human Rights, which provide a framework for analyzing the responsibility of private-sector actors involved in the development and deployment of AI systems. This is necessary because many contemporary privacy interferences are not carried out directly by States, but by private companies whose systems structure communication, emotional support, information access, and behavioral environments.

To bridge the gap between traditional legal categories and ai-mediated harms, the thesis also draws upon theoretical scholarship on intellectual privacy, mental autonomy, chilling effects, and the structural effects of AI systems. This scholarship provides the conceptual vocabulary necessary to articulate harms that current jurisprudence may not fully capture, including environmental modulation, anticipatory conformity, behavioral modification,

¹⁷ *ibid.*, p. 218.

and “slow violence”.¹⁸ These concepts are not used to replace legal doctrine, but to clarify how existing standards of proportionality and positive obligations should be applied to cumulative, psychological, and structurally embedded harms.

The primary case studies used to test the limits of existing proportionality doctrine are *NJCM c.s. and others v De Staat der Nederlanden* (SyRI) and *Bărbulescu v Romania*. These cases are selected because they offer particularly relevant examples of judicial engagement with algorithmic risk assessment and workplace monitoring. SyRI illustrates the difficulties of legality, transparency, and proportionality in the context of state-led data processing and risk prediction. *Bărbulescu* illustrates how positive obligations may require States to secure respect for private life within relationships involving private actors. Together, these cases demonstrate both the capacity of existing doctrine and the emergence of a proportionality gap where cumulative psychological and structural harms remain insufficiently recognized.

1.3 Scope and Limitations

The scope of this thesis is limited to the right to privacy as protected under Article 17 of the International Covenant on Civil and Political Rights and Article 8 of the European Convention on Human Rights. While artificial intelligence systems implicate a broad range of human rights, related concepts such as freedom of thought, human dignity, data protection, non-discrimination, and business and human rights are addressed only where they are directly relevant to the privacy argument. In particular, references to mental autonomy and the inner sphere are used to clarify how privacy protects the conditions necessary for autonomous self-development, rather than to provide a full analysis of freedom of thought under Article 18 ICCPR.

The thesis does not provide a technical or computer-science analysis of AI architectures. Nor does it attempt to offer a comprehensive study of all AI-related human rights issues. Instead, it focuses on the human rights implications of AI-mediated inference, prediction, and environmental modulation. The analysis is therefore concerned with how AI systems may affect autonomy, identity formation, and the individual’s inner sphere.

The thesis is specifically oriented toward structural and psychological privacy harms. These include chilling effects, anticipatory conformity, behavioral modification and emotional dependency. More visible harms, such as financial loss, physical detention, or direct discrimination, are considered only where they assist the central argument concerning privacy and the inner sphere.

1.4 Outline

¹⁸ Sue Anne Teo. Artificial intelligence and its ‘slow violence’ to human rights. *AI Ethics* 5, (2025)

The thesis consists of following chapters, chapter 1 introduces the background, research objectives, research questions, methodology, scope, and structure of the thesis. It presents the central problem: the challenge posed by AI is not necessarily the absence of privacy doctrine, but the under-inclusive application of proportionality analysis to cumulative and psychologically mediated harms.

Chapter 2 develops the doctrinal foundations of privacy under international human rights law. It examines Article 17 ICCPR and Article 8 ECHR, focusing on legality, foreseeability, necessity, proportionality, safeguards, and positive obligations. It then argues that privacy should be understood not only as protection against informational intrusion, but also as a safeguard for the individual's inner sphere and autonomous self-development.

Chapter 3 analyses the specific harms generated by AI systems. It examines how behavioral inference, digital phenotyping, predictive analytics, emotional inference, and environmental modulation transform privacy interferences into continuous and often invisible processes of observation and influence. The chapter further considers chilling effects and algorithmic vulnerability, particularly in relation to children, young people, and adults in crisis.

Chapter 4 evaluates state responsibility through the case studies of *SyRI* and *Bărbulescu*. It argues that while existing doctrine contains important mechanisms for addressing AI-mediated privacy interferences, judicial and regulatory analysis often remains attached to a certain harm model. The chapter also examines emerging AI governance frameworks, including the EU AI Act and human rights due diligence, assessing whether they are capable of addressing structural and psychological harms rather than merely procedural compliance.

Chapter 5 concludes the thesis by arguing international human rights law should recognize cumulative psychological and structurally embedded harms as legally cognizable interferences with the private sphere. This does not require abandoning existing doctrine, but applying it in a way that protects the conditions under which individuals think, choose, develop identity, and exercise autonomy.

2 Privacy, Proportionality, and the “Inner Sphere” Under International Human Rights Law

2.1 The Right to Privacy

The human mind forms the foundation of thought, autonomy, identity, and decision-making. Every action, belief, preference, and expression originates internally before manifesting externally. In this sense, the protection of the individual’s inner mental sphere is essential to the meaningful exercise of human freedom itself.

This raises a deeper question concerning the relationship between the human mind and the right to privacy. Privacy itself remains a broad and evolving concept within international human rights law (IHRL), not susceptible to exhaustive definition. As Neil Richards argues through the concept of “intellectual privacy”,¹⁹ privacy functions as a “zone of protection” within which individuals are able to think, explore ideas, develop beliefs, and form identity free from excessive observation or interference.²⁰ Within this protected sphere, privacy becomes more than a safeguard for information alone; it becomes a condition necessary for independent thought and autonomous self-development.

The right to privacy therefore constitutes a foundational guarantee within IHRL, protecting individuals against arbitrary or unlawful interference with their private sphere. Article 17 of the International Covenant on Civil and Political Rights provides that:

“No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence.”

The ICCPR framework, as interpreted by the Human Rights Committee (HRC) and legal scholars such as Joseph and Castan²¹, draws a strict distinction between “unlawful” and “arbitrary” interferences with privacy. This distinction becomes particularly important because Article 17 does not merely prohibit illegal interferences but seeks to constrain the broader exercise of State power over the individual’s private sphere.

An interference is considered “unlawful” unless it is expressly envisaged by domestic legislation that itself complies with the provisions, aims, and objectives of the Covenant.²² Importantly, as emphasized by the HRC, the mere existence of legislation alone does not automatically legitimize an interference with privacy. Rather, the “quality of law” itself becomes central

¹⁹ Neil M Richards, ‘Intellectual Privacy’ (2008) 87 *Texas Law Review* 387.

²⁰ Neil M Richards, *Intellectual Privacy: Rethinking Civil Liberties in the Digital Age* (Oxford University Press 2015) 95.

²¹ Sarah Joseph and Melissa Castan, *The International Covenant on Civil and Political Rights: Cases, Materials, and Commentary* (3rd edn, OUP 2013) 537.

²² *ibid.*, p. 535.

to the protection afforded under Article 17. Legislation authorizing surveillance, interception, data collection, or other forms of interference must therefore remain sufficiently precise and circumscribed, specifying in detail the exact circumstances under which interferences are permitted.²³ As Joseph and Castan observe, this requirement functions as an important safeguard against excessive concentrations of State power by ensuring that individuals remain capable of foreseeing the circumstances in which their privacy may be interfered with. The requirement of legal precision therefore becomes particularly significant within technologically enhanced systems of governance where broad or vague legal mandates may permit extensive forms of algorithmic surveillance and data processing without meaningful limitation or accountability. In this regard, the HRC stresses that privacy protections are intended to prevent authorities from exercising “unfettered discretion,” which remains inherently incompatible with the right to privacy itself.²⁴

Even where an interference is formally “lawful” under municipal rules, it may nevertheless violate Article 17 where it is “arbitrary.”²⁵ The concept of arbitrariness therefore extends beyond illegality alone and introduces a substantive reasonableness requirement into the protection of privacy rights. As clarified by the HRC, arbitrariness ensures that every form of State action — whether legislative, administrative, or judicial — remains reasonable and proportionate to the objective pursued in the specific circumstances of the case.²⁶

This distinction becomes particularly important because it prevents States from relying solely upon procedural legality to justify intrusive forms of surveillance or technological governance. In other words, State authorities cannot simply enact broad legislative powers and assume compliance with Article 17; rather, the interference itself must remain justified, necessary, and proportionate in substance. The concept of arbitrariness therefore operates as an important safeguard against formally lawful yet substantively excessive interferences with private life.

As established in *Toonen v Australia*, the principle of reasonableness further requires that an interference constitute the least intrusive instrument available to achieve a legitimate aim and remain proportionate to the interest protected. Authorities may therefore only call for information that is genuinely “essential in the interests of society.”²⁷ This proportionality framework becomes increasingly significant within artificial intelligence mediated environments where contemporary technologies permit unprecedented forms of behavioral inference, predictive profiling, and large-scale algorithmic surveillance. Consequently, proportionality analysis within the context of artificial intelligence (AI) cannot remain limited solely to visible or informational intrusions but must also account for the broader

²³ *ibid.*, p. 536.

²⁴ Human Rights Committee, *General Comment No 16: Article 17 (Right to Privacy)* (8 April 1988) para. 11.

²⁵ Joseph and Castan (n 13) p. 537

²⁶ *ibid.*, p. 537

²⁷ *ibid.*

implications such systems may have upon autonomy, behavioral freedom, and the individual's inner sphere of self-development.

Under the ECHR Article 8 serves as a critical mechanism for balancing individual privacy against the interests of the state, permitting interference only where such measures are “in accordance with the law” and “necessary in a democratic society.”²⁸ This two-pronged framework becomes particularly important within contemporary systems of secret surveillance and technologically enhanced governance, where expanding State capabilities increasingly risk undermining the effective protection of private life. In this regard, Article 8 functions not merely as a formal procedural safeguard, but as a substantive limitation upon excessive or arbitrary exercises of State power. As the European Court of Human Rights (ECtHR) has repeatedly emphasized, interferences with private life must remain governed by the rule of law and subject to rigorous European supervision, particularly where surveillance technologies permit extensive forms of monitoring, profiling, and data analysis at unprecedented scale.²⁹

2.1.1 The Legality Requirement

For an interference to be considered “in accordance with the law,” it must possess a specific basis within domestic legislation and remain compatible with the broader principles of the rule of law.³⁰ Importantly, the ECtHR has consistently clarified that this requirement extends beyond the mere existence of a statutory provision authorizing surveillance or data collection. Rather, Article 8 incorporates a broader “quality of law” standard requiring that the legal framework itself remain sufficiently accessible, precise, and foreseeable as to its effects.³¹ Individuals must therefore be capable of understanding, at least to a reasonable degree, the circumstances under which public authorities may interfere with their private life.³²

Within the context of secret surveillance, however, the Court recognizes that “foreseeability” cannot require individuals to predict the exact moment at which authorities may monitor them, since such a requirement would undermine the operational effectiveness of covert surveillance measures altogether.³³ Instead, the ECtHR requires that the law provide citizens with an adequate indication of the scope of discretion exercised by public authorities, as well as the conditions and procedures governing the use of intrusive surveillance powers.³⁴ The underlying concern here is

²⁸ *Big Brother Watch and Others v United Kingdom* App Nos 58170/13, 62322/14 and 24960/15 (ECtHR, 25 May 2021), para. 273

²⁹ *ibid.*, paras. 338-339.

³⁰ *ibid.*, paras. 331-332.

³¹ Zuzanna Warso, ‘Human Rights Requirements for Person-Based Predictive Policing: Lessons from Selected ECtHR Case Law and Its Limits’ (2022) *Technology and Regulation*, p. 75.

³² *ibid.*, p. 76.

³³ *Big Brother Watch*, (n 28), para. 333.

³⁴ *ibid.*

fundamentally tied to the prevention of arbitrary power. Where surveillance regimes are framed in excessively broad, vague, or opaque terms, individuals become unable to meaningfully foresee the extent to which their communications, behavior, or personal data may become subject to State scrutiny.

This issue becomes particularly significant within AI-mediated systems where contemporary technologies increasingly permit continuous data aggregation, predictive profiling, behavioral inference, and automated analysis of personal information. In such contexts, vague legislative authorizations may permit forms of technologically enhanced surveillance extending far beyond what individuals can reasonably anticipate or contest. Consequently, the “quality of law” requirement operates as a critical safeguard against technologically facilitated arbitrariness by ensuring that surveillance powers remain constrained by clear legal standards rather than unchecked executive discretion.³⁵

2.1.2 Heightened Standards for Technology

As surveillance technologies become increasingly sophisticated, the ECtHR has acknowledged that the corresponding risks of abuse, opacity, and excessive executive discretion likewise intensify. The Court therefore demands “particularly precise” legal rules governing technologically advanced surveillance regimes.³⁶ This heightened standard reflects the recognition that contemporary digital systems are capable not merely of collecting isolated pieces of information, but of generating extensive behavioral insights, relational mapping, and predictive inferences concerning individuals’ lives.

Accordingly, the “quality of law” requirement under Article 8 necessarily implies the existence of effective safeguards against arbitrary interference at every stage of the surveillance process.³⁷ In *Big Brother Watch and Others v The United Kingdom*, the Grand Chamber emphasized that any bulk surveillance regime must contain comprehensive “end-to-end safeguards” in order to remain compatible with the Convention.³⁸ Importantly, these safeguards must operate throughout the entire surveillance cycle, from the initial interception and retention of communications to the application of selectors, filtering procedures, algorithmic processing, and eventual human examination of collected material.³⁹

The Court therefore stressed that domestic law must clearly define both the procedures governing surveillance operations and the forms of independent oversight exercised over them. In particular, the ECtHR

³⁵ Zuzanna Warso, ‘Human Rights Requirements for Person-Based Predictive Policing: Lessons from Selected ECtHR Case Law and Its Limits’ (2022) *Technology and Regulation*, p. 76.

³⁶ *Big Brother Watch*, (n 28), para. 333.

³⁷ *ibid.*, para. 334.

³⁸ *ibid.*, para. 350.

³⁹ *ibid.*, para. 356.

identified six minimum safeguards that must be expressly set out within domestic legislation:⁴⁰

- the nature of offences capable of giving rise to interception orders;
- the categories of individuals liable to have communications intercepted;
- limitations upon the duration of surveillance measures;
- procedures governing the examination, use, and storage of collected data;
- precautions regulating the communication of data to third parties; and
- the circumstances under which data must be erased or destroyed.

These safeguards become especially important in the context of AI-driven surveillance systems because automated technologies increasingly permit large-scale analysis, cross-platform aggregation, behavioral prediction, and inferential profiling in ways that are often invisible to affected individuals themselves.

2.1.3 The Proportionality Test

The second prong of the Article 8 framework requires that any interference with private life be “necessary in a democratic society” for the achievement of a legitimate aim, including national security, public safety, or the prevention of crime.⁴¹ To satisfy this requirement, States must demonstrate that the interference corresponds to a “pressing social need” and remains proportionate to the legitimate objective pursued.⁴²

The proportionality assessment therefore requires a balancing exercise in which the seriousness of the intrusion into the individual’s private life is weighed against the operational or investigative interests relied upon by the State.⁴³ Importantly, the ECtHR has consistently emphasized that surveillance measures must remain the least intrusive instrument available for achieving the desired result. Where information may reasonably be obtained through less intrusive means, the interference will not satisfy the proportionality requirement.⁴⁴

The ECtHR has additionally recognized that States claiming a “pioneer role” in the development and deployment of new surveillance technologies bear a “special responsibility” to strike an appropriate balance between technological innovation and the protection of private life. The potential utility of advanced scientific techniques cannot justify weakening the protection afforded under Article 8 at any cost.⁴⁵ Consequently, the Court’s assessment of necessity ultimately depends upon the existence of adequate

⁴⁰ *ibid.*, para. 335

⁴¹ *ibid.*, para. 273

⁴² *ibid.*, paras. 331-332.

⁴³ Warso, (n 35) p. 78.

⁴⁴ *Big Brother Watch*, (n 28), para. 3.7.

⁴⁵ Warso, (n 35) p. 78.

and effective guarantees against abuse, including independent authorization procedures, meaningful oversight mechanisms, and robust ex post facto review conducted by judicial or independent bodies.⁴⁶ Without such safeguards, technologically enhanced surveillance regimes fail both the “quality of law” requirement and the broader proportionality standard necessary within a democratic society.⁴⁷

2.2 Privacy Beyond Informational Secrecy

Modern legal scholarship increasingly recognizes that privacy is not merely an individual preference or a defensive claim against intrusion, but rather the “matrix” and “indispensable condition” for the meaningful exercise of nearly all other forms of freedom.⁴⁸ Within this broader understanding, Sarah Joseph and Melissa Castan, building upon the work of Manfred Nowak, argue that the right to privacy under Article 17 ICCPR encompasses freedom from unwarranted and unreasonable intrusions into activities that belong to the realm of individual autonomy.⁴⁹ Importantly, this realm of autonomy refers to a sphere of action that does the liberty of others and within which individuals may withdraw in order to shape their lives according to their own wishes, expectations, and sense of self.⁵⁰ In this regard, privacy functions not merely as a protection for secrecy or informational control, but as a safeguard for the conditions necessary for autonomous personality development itself.⁵¹

Similarly, the HRC clarifies that privacy refers to the sphere of a person’s life within which they may freely express their identity, whether alone or in relationships with others.⁵² This doctrinal understanding becomes particularly significant because it establishes what may be described as an “inner personal core” essential for the construction of the self and the development of personality.⁵³ As Nowak further observes, the notion of identity protected under the ICCPR extends broadly beyond external identifiers alone and includes a person’s name, appearance, feelings, and thoughts.⁵⁴ Consequently, the protection of privacy under Article 17 increasingly appears inseparable from the protection of the internal processes through which individuals form identity, develop beliefs, and cultivate emotional and intellectual independence.

This inner sphere becomes especially important because it is within this protected mental and emotional space that individuals engage in the formative processes of self-differentiation, gradually distinguishing themselves from

⁴⁶ *Big Brother Watch*, (n 10), para. 339.

⁴⁷ *ibid.*, p. 353.

⁴⁸ Richards, (n 19), pp. 389-390.

⁴⁹ Joseph and Castan, (n 13) pp. 533-534.

⁵⁰ *ibid.*, p. 534.

⁵¹ *ibid.*, p. 534 [16.03]

⁵² *ibid.*

⁵³ Zuzanna Istace, ‘Protecting the Mental Realm: What Does Human Rights Law Bring to the Table?’ (2023) 41(4) *Netherlands Quarterly of Human Rights*, pp. 231-232.

⁵⁴ *ibid.*

surrounding social, political, and cultural pressures.⁵⁵ Without meaningful legal safeguards protecting these internal dimensions of human life, including emotions, imagination, intentions, and intellectual exploration, individuals risk becoming vulnerable to conformity pressures, behavioral conditioning, and forms of self-censorship capable of undermining authentic self-development.⁵⁶ This concern becomes even more pronounced within AI-mediated environments where algorithmic systems increasingly possess the capacity not merely to observe behavior, but to shape informational environments, behavioral tendencies, emotional responses, and processes of decision-making themselves.

Julie Cohen further advances this understanding of privacy by conceptualizing it as “breathing room” necessary for socially situated individuals to engage in processes of boundary management and self-formation.⁵⁷ Importantly, Cohen rejects the idea of the self as a “precultural island” existing independently from society.⁵⁸ Rather, she argues that subjectivity emerges through a continuous interaction between social shaping and individual agency.⁵⁹ Within this framework, privacy becomes a dynamic structural condition preventing identity formation and communal values from developing entirely in “lockstep” with dominant institutional or technological systems.⁶⁰ By preserving interstitial spaces within broader social and technical structures, privacy enables individuals to engage in self-articulation, experimentation, and self-determination. In other words, privacy creates the necessary conditions for the ongoing process of “self-making,” allowing individuals to navigate social and cultural influences without becoming wholly absorbed by them.⁶¹

Neil Richards complements these theories through the concept of “intellectual privacy,” which he describes as a vital precondition for freedom of thought and expression.⁶² Richards argues that traditional free speech theory often focuses excessively upon the “marketplace of ideas” while neglecting the more intimate “workshops” within which ideas are actually developed.⁶³ Intellectual privacy therefore protects the internal processes of thinking, reading, exploring ideas, and generating beliefs away from unwanted observation or interference. According to Richards, this protection contains three interrelated dimensions: freedom of thought and belief, the right to read and engage in intellectual exploration without fear of scrutiny, and the confidentiality of communications allowing individuals to test ideas privately before expressing them publicly. Ultimately, intellectual privacy

⁵⁵ Julie E Cohen, ‘What Privacy Is For’ (2013) 126 *Harvard Law Review* pp. 1905-1908 <<https://harvardlawreview.org/print/vol-126/what-privacy-is-for/>> accessed 21 May 2026.

⁵⁶ *Istace*, (n 53) p. 215.

⁵⁷ Julie E Cohen, ‘Surveillance vs Privacy: Effects and Implications’ in David Gray and Stephen E Henderson (eds), *Cambridge Handbook of Surveillance Law* (CUP 2017) p. 5.

⁵⁸ *ibid.*

⁵⁹ *ibid.*

⁶⁰ Cohen, (n 55) p. 1911.

⁶¹ *ibid.*

⁶² Richards, (n 19), pp. 388-389.

⁶³ *ibid.*, pp. 395-396.

safeguards what Richards describes as the “engine of expression” itself, the human mind.⁶⁴ Consequently, creating conditions under which cognition itself risks becoming distorted by the persistent awareness of observation and algorithmic evaluation.

2.3 The “Inner Sphere”

Under IHRL, the “inner sphere” or *forum internum* has increasingly transitioned from what was once presumed to be an “impregnable fortress” into a site of profound legal vulnerability.⁶⁵ Historically, the human mind was largely understood as existing under the exclusive control of the individual and therefore beyond meaningful external access or intrusion.⁶⁶ However, contemporary technological developments increasingly challenge this foundational assumption. Emerging technologies such as digital phenotyping, predictive analytics, affective computing, and conversational AI systems are now capable of inferring, analyzing, and responding to deeply personal aspects of human cognition and behavior in ways that significantly narrow the practical distance between external observation and internal mental life. As a result, the long standing presumption that mental processes remain naturally inaccessible to external actors becomes increasingly difficult to sustain within digitally mediated environments.⁶⁷

Importantly, this inner sphere encompasses far more than abstract thought alone. It includes an individual’s emotions, imagination, intentions, behavioral tendencies, and decision making processes.⁶⁸ In this regard, the protection of the *forum internum* becomes essential not merely for preserving secrecy, but for safeguarding the conditions necessary for autonomous self construction and authentic participation within society itself.⁶⁹ The ability to think, deliberate, imagine, question, and form identity free from excessive manipulation or observation remains central to the meaningful exercise of human freedom and agency.⁷⁰ Consequently, contemporary threats to mental privacy increasingly implicate broader concerns relating to dignity, authenticity, and the preservation of independent selfhood.

Within this evolving discourse, scholars increasingly distinguish between mental privacy and mental autonomy. Mental privacy refers to the individual’s ability to control when, how, and to what extent their thoughts, emotions, or internal mental states become externally accessible or communicable. Mental autonomy, by contrast, concerns the capacity to independently control one’s own mental functions, including episodic

⁶⁴ *ibid.*, pp. 403-404.

⁶⁵ Istace, (n 53) pp. 214-215.

⁶⁶ *ibid.*

⁶⁷ Lisa Cosgrove, Justin M Karter, Mallaigh McGinley and Zenobia Morrill, ‘Digital Phenotyping and Digital Psychotropic Drugs: Mental Health Surveillance Tools That Threaten Human Rights’ (2020) 22(2) *Health and Human Rights Journal*, p. 33.

⁶⁸ Istace, (n 53) p. 215.

⁶⁹ *ibid.*

⁷⁰ *ibid.* p. 216.

memory, rational deliberation, planning, emotional processing, and decision making.⁷¹ This distinction becomes particularly important in the context of AI driven systems because technologically mediated interference may occur even where no direct disclosure of thoughts takes place. When algorithmic systems continuously shape informational environments, behavioral options, emotional responses, or cognitive patterns through recommendation systems, predictive profiling, and personalized engagement mechanisms, the resulting interference extends beyond the extraction of information alone. In such circumstances, the concern is not merely that data about the individual becomes known to third parties, but that the individual's own processes of autonomous thought formation and decision making may themselves become subtly modulated, redirected, or constrained. Consequently, such interferences increasingly threaten the authenticity, integrity, and agency of personhood itself.⁷²

One of the most significant harms emerging within this inner sphere is the phenomenon commonly described as the “chilling effect.”⁷³ Jonathon Penney argues that chilling effects should not be understood merely as isolated acts of self-censorship motivated by fear of legal punishment. Rather, Penney conceptualizes chilling effects as a form of compelled social conformity arising from conditions of uncertainty and perceived observation. Unlike traditional deterrence theories that focus primarily upon explicit sanctions, Penney's social conformity framework suggests that individuals who become aware that they may be monitored gradually adapt their behavior toward perceived social expectations in order to avoid disapproval, suspicion, or social exclusion. In this sense, the harm generated by pervasive surveillance becomes structural and psychological rather than purely coercive.⁷⁴

This dynamic is closely reflected in the “panopticon effect,” whereby the mere possibility of observation induces forms of anticipatory self-discipline and “normalized soul training.” Importantly, the effectiveness of such systems does not necessarily depend upon continuous active monitoring. Rather, the awareness or perception that one may be watched becomes sufficient to encourage behavioral adaptation and internalized conformity.⁷⁵ Within AI mediated environments, this effect becomes particularly significant because algorithmic systems are increasingly embedded within everyday communicative, social, and informational infrastructures. As a result, individuals may gradually regulate their speech, associations, preferences, and patterns of behavior in response to persistent forms of invisible observation and evaluation.

Pervasive surveillance therefore risks eroding mental autonomy by effectively “short circuiting” independent decision-making processes and subtly inclining individuals toward behavior that is socially acceptable,

⁷¹ *ibid.*

⁷² Cohen, (n 55) *Harvard Law Review* p. 1912.

⁷³ Penney, (n 10) p. 1455.

⁷⁴ *ibid.*, p. 1488.

⁷⁵ *ibid.*, p. 1491.

predictable, and normatively “safe.”⁷⁶ In this regard, surveillance does not merely observe behavior after it occurs, but increasingly shapes the conditions under which behavior is formed in the first place. The resulting pressure toward the “bland and the mainstream” threatens not only freedom of expression, but the broader human capacity for independent thought, experimentation, dissent, and authentic self development.⁷⁷

2.4 Proportionality Under International Human Rights Law

The proportionality test operates as the primary mechanism through which courts determine whether an interference with the right to privacy may be justified under IHRL. Under both Article 17 ICCPR and Article 8 ECHR, any restrictive measure must remain reasonable, necessary in a democratic society, and constitute the least intrusive instrument available for achieving a legitimate aim.⁷⁸ In this regard, proportionality functions as a critical safeguard against excessive or arbitrary exercises of State power by requiring that interferences with privacy remain carefully balanced against the objective pursued. States must therefore demonstrate not only that the interference pursues a legitimate aim, such as national security, public safety, or public health, but also that the size, intensity, and scope of the interference remain proportionate to what is sought to be achieved.⁷⁹ The proportionality framework thus attempts to ensure that privacy restrictions do not exceed what is genuinely necessary within a democratic society.

However, a significant “proportionality gap” increasingly emerges within cases involving mass surveillance, predictive technologies, and experimental algorithmic systems. While proportionality doctrine formally requires rigorous balancing and evidentiary justification, courts frequently accept broad and highly abstract governmental aims, particularly “national security,” as sufficient grounds for intrusive surveillance measures without demanding meaningful empirical evidence demonstrating the actual effectiveness, reliability, or necessity of the technology deployed.⁸⁰ As Zuzanna Warso observes, this creates a profound “blind spot” within legal review itself. If a technological system cannot be independently verified as accurate, effective, or operationally necessary, then its deployment cannot logically satisfy the strict necessity requirement underlying proportionality analysis.⁸¹ In this sense, proportionality risks becoming overly deferential precisely in contexts

⁷⁶ Cohen, (n 57) p. 7.

⁷⁷ *ibid.*, p. 9.

⁷⁸ Human Rights Committee, *General Comment No 27: Freedom of Movement* (1999) UN Doc CCPR/C/21/Rev.1/Add.9, para 14; *Handyside v United Kingdom* (1976) 1 EHRR 737, paras. 14-15.

⁷⁹ *Big Brother Watch*, (n 28)

⁸⁰ Warso, (n 35) p. 77.

⁸¹ *ibid.*, p. 79.

where technologically enhanced State powers possess the greatest capacity for intrusion and abuse.

This difficulty becomes even more pronounced within AI mediated environments because contemporary technologies increasingly generate harms that are diffuse, cumulative, psychological, and structurally embedded rather than immediate or directly visible. Existing judicial doctrine often struggles to adequately recognize harms such as pervasive anxiety, anticipatory conformity, behavioral adaptation, and self-regulation induced by conditions of constant observation.⁸² These harms rarely manifest as singular or easily measurable injuries. Instead, they emerge gradually through prolonged exposure to surveillance environments and persistent awareness of potential monitoring.⁸³

The limitations of existing doctrine become particularly visible in *Clapper v Amnesty International USA*, where the United States Supreme Court dismissed chilling effects as “self-inflicted injuries” arising merely from “subjective fear,” considering them too speculative to establish standing. Importantly, this reasoning reflects a narrow understanding of surveillance harms that focuses primarily upon direct material injury while overlooking the broader social and psychological consequences of pervasive observation.⁸⁴ Such an approach fails to fully account for the reality that chilling effects are not simply internal emotional reactions, but externally produced behavioral outcomes generated through systemic observation, social conformity pressures, and uncertainty regarding surveillance exposure.⁸⁵

Contemporary empirical research increasingly substantiates these concerns by demonstrating that awareness of monitoring may significantly alter patterns of communication, association, intellectual exploration, and political participation. Consequently, the proportionality framework faces growing difficulty in adequately addressing forms of harm that do not arise through immediate physical intrusion or direct disclosure of information, but through gradual modifications of behavior, cognition, and autonomous self-development.⁸⁶ Within AI driven systems capable of continuous behavioral inference and predictive evaluation, these limitations become particularly acute because the resulting interference extends beyond traditional informational privacy and increasingly implicates the integrity of the individual’s inner sphere itself.

2.5 State Obligations, Private Actors, and the UN Guiding Principles

⁸² Penney, (n 10) p. 1528.

⁸³ *ibid.*

⁸⁴ *ibid.*, p. 1460.

⁸⁵ *ibid.*, p. 1528.

⁸⁶ *ibid.*, p. 1508.

The transition of the private sphere from an “impregnable fortress” to a site of intense legal vulnerability increasingly necessitates a robust framework of State positive obligations and corporate responsibility aimed at protecting the forum internum.⁸⁷ Under contemporary conditions of AI driven governance and “surveillance capitalism,” the protection of privacy can no longer be understood solely through a negative obligation framework requiring States to refrain from arbitrary interference. Rather, IHRL increasingly recognises that States possess affirmative obligations to actively protect individuals against harmful interferences originating from private sector actors, particularly where technologically mediated systems possess the capacity to affect mental autonomy, behavioural freedom, and the integrity of the individual’s inner sphere.

2.5.1 The Doctrinal Basis of Positive Obligations

The doctrinal basis for these positive obligations emerges clearly within the Human Rights Committee’s interpretation of Article 17 ICCPR. In General Comment No. 16, the HRC clarifies that the right to privacy must be protected against interferences regardless of whether they emanate from State authorities or from natural or legal persons, including private corporations. Consequently, the protection of privacy under Article 17 extends beyond preventing direct State intrusion and requires States to adopt legislative, administrative, and regulatory measures governing the gathering, storage, and use of personal information by private entities.⁸⁸

Under Article 8 ECHR, this obligation is similarly framed through the doctrine of positive obligations. The State’s responsibility may therefore become engaged not only where it directly interferes with privacy rights, but also where it fails to adequately secure those rights within relationships between private parties, including employers, digital platforms, and technology companies.⁸⁹ This doctrinal development becomes increasingly important in the context of AI systems because contemporary digital infrastructures often permit extensive forms of behavioural monitoring, emotional inference, algorithmic recommendation, and predictive profiling carried out by private actors rather than States directly. In such circumstances, the absence of adequate regulatory safeguards may itself constitute a failure by the State to effectively protect the right to private life.

2.5.2 The UN Guiding Principles

The UN Guiding Principles (UNGPs) on Business and Human Rights provide a particularly important framework for addressing these emerging

⁸⁷ Cohen, (n 57) p. 6.

⁸⁸ Office of the United Nations High Commissioner for Human Rights, *Guiding Principles on Business and Human Rights: Implementing the United Nations “Protect, Respect and Remedy” Framework* (United Nations 2011)

⁸⁹ Joseph and Castan, (n 13) p. 540

forms of private sector interference within AI mediated environments. Through the “Protect, Respect and Remedy” framework, the UNGPs establish a structured model for allocating responsibility between States and corporate actors in situations involving human rights risks arising from business activity.⁹⁰

2.5.2.1 The State Duty to Protect

Under Pillar I, States must protect individuals against human rights abuses committed by third parties through effective policies, legislation, regulation, and adjudication.⁹¹ In the context of AI technologies, this obligation increasingly requires States to ensure that existing privacy and data protection frameworks remain capable of addressing harms generated by evolving technological systems. Importantly, this obligation cannot remain static. States must continuously assess whether legal frameworks remain adequate considering increasingly sophisticated forms of data extraction, behavioral inference, emotional profiling, and algorithmic manipulation.

This obligation becomes particularly significant where States increasingly rely upon or privatize digital infrastructures affecting fundamental rights, including healthcare, therapeutic support, education, and communication platforms.⁹² States do not relinquish their human rights obligations merely because harmful practices originate within private technological systems rather than traditional State institutions.⁹³ Consequently, failures to regulate AI driven systems capable of significantly interfering with mental autonomy or behavioral freedom may engage the State’s positive obligations under both Article 17 ICCPR and Article 8 ECHR.

2.5.2.2 The Corporate Responsibility to Respect

Business enterprises possess an independent responsibility to respect human rights regardless of whether States fully discharge their own obligations. Importantly, this responsibility exists as a global standard of expected conduct and does not depend entirely upon domestic enforcement mechanisms.⁹⁴ Companies must therefore avoid infringing upon human rights and address adverse impacts with which they are involved.

Under the UNGP framework, such influence necessarily carries corresponding responsibilities relating to privacy, dignity, autonomy, and mental integrity.

A central mechanism for operationalizing this responsibility is Human Rights Due Diligence (HRDD). Principles 17 to 21 require companies to identify, prevent, mitigate, and account for adverse human rights impacts through an ongoing due diligence process. In the context of AI systems, this

⁹⁰ OHCHR, *Guiding Principles on Business and Human Rights* (n 88)

⁹¹ *ibid.*, p. 3.

⁹² *ibid.*

⁹³ *ibid.*, p. 8.

⁹⁴ *ibid.*, p. 13.

requires companies to assess not only technical risks such as data breaches or discriminatory outputs, but also broader psychological and relational harms arising from algorithmic engagement systems. Importantly, AI developers must consider the specific context within which technologies operate, including highly sensitive environments such as therapeutic, emotional support, or mental health related interactions. Particular attention must additionally be paid to individuals or groups experiencing heightened vulnerability, dependency, or marginalization.⁹⁵

The issue of “leverage” further becomes significant where companies contribute to harmful impacts through algorithmic systems capable of eroding mental autonomy or shaping behavioral environments. Under the UNGPs, where a company contributes to a harmful impact, it possesses a responsibility not only to cease that contribution but also to utilise its leverage to mitigate continuing harms. Within AI mediated environments, this may include altering engagement architectures, limiting manipulative recommendation systems, restricting exploitative emotional modelling practices, or redesigning systems that incentivize compulsive dependency and behavioral conditioning.⁹⁶

Although the primary focus of this thesis is the State’s duty to protect, the corporate responsibility to respect is relevant because it identifies the human rights risks that States must require companies to identify, prevent, mitigate, and account for through regulation and oversight.

2.5.2.3 *Access to Remedy*

States must ensure that individuals affected by human rights abuses possess access to effective judicial and non-judicial remedies.⁹⁷ However, the implementation of meaningful remedies becomes particularly difficult within AI driven environments due to the “black box” nature of algorithmic systems. Individuals frequently remain unable to understand how decisions are generated, how behavioral profiles are constructed, or how algorithmic outputs affect their opportunities, emotions, or decision-making processes.⁹⁸

Consequently, effective remedies increasingly require States to reduce both legal and practical barriers to accountability. This includes ensuring transparency obligations, independent oversight, accessible complaint mechanisms, and meaningful avenues through which individuals may contest algorithmic harms.⁹⁹ Without such safeguards, the ability to challenge technologically mediated interferences with privacy risks becoming largely illusory.

The remedy section is relevant to this thesis only in a limited sense. It is not examined as an independent corporate accountability framework, but as

⁹⁵ *ibid.*, pp. 19-20.

⁹⁶ *ibid.*, pp. 20-21.

⁹⁷ *ibid.*, p. 28.

⁹⁸ *ibid.*, p. 29; Julie E Cohen, ‘What Privacy Is For’ (2013) 126 *Harvard Law Review* p. 1914.

⁹⁹ OHCHR, *Guiding Principles on Business and Human Rights* (n 88) p. 29.

part of the State's positive obligation to ensure that individuals affected by AI-mediated privacy interferences have meaningful avenues to challenge opaque, inferential, or psychologically harmful systems.

3 AI, Privacy, and the Structural Erosion of the Private Sphere

The deployment of sophisticated AI systems necessitates a significant re-evaluation of privacy interferences under IHRL. Whereas traditional doctrinal protections under Article 17 ICCPR and Article 8 ECHR primarily focused upon preventing external access to discrete personal information, contemporary AI systems increasingly shift the locus of interference toward behavioral inference, environmental modulation, and the gradual restructuring of the conditions under which individuals think, decide, and develop identity.¹⁰⁰ In this sense, AI systems do not merely collect data; they increasingly shape the informational and behavioral environments within which autonomy is exercised.¹⁰¹

This chapter identifies three interrelated forms of AI-mediated privacy harm. The first is inferential harm, where AI systems derive intimate conclusions about emotions, vulnerabilities, preferences, and future conduct from seemingly ordinary digital traces. The second is modulatory harm, where algorithmic systems shape the informational and behavioral environment in which individuals think, choose, and act. The third is vulnerability-based harm, where AI systems intensify existing vulnerabilities or generate new ones through dependency, opacity, emotional targeting, and asymmetries of knowledge and power. Together, these harms demonstrate why privacy interferences in the AI context cannot be assessed only through visible acts of data collection or disclosure, but must also be evaluated through their cumulative effects on autonomy, identity formation, and the individual's inner sphere.

A central theoretical contribution informing this analysis is the work of Sue Anne Teo, who argues that AI-driven systems facilitate the granular targeting and tailoring of choice architectures to the particular characteristics, vulnerabilities, and behavioral tendencies of individual users.¹⁰² Teo's framework is particularly relevant because it moves beyond the traditional "violations vernacular" of informational secrecy and instead focuses upon the broader "conditions of possibility" necessary for meaningful autonomy and human self-development.¹⁰³ Her analysis therefore provides an important conceptual foundation for understanding how AI systems may interfere with human rights even in the absence of direct coercion or visible intrusion.

However, this thesis remains analytically and doctrinally distinct from Teo's project in legal scope. Whereas Teo primarily seeks to revitalize the protection of freedom of thought under Article 18 ICCPR, this thesis

¹⁰⁰ United Nations Human Rights Council, *The Right to Privacy in the Digital Age: Report of the United Nations High Commissioner for Human Rights* UN Doc A/HRC/39/29 (3 August 2018) para. 15.

¹⁰¹ S A Teo, 'How to think about freedom of thought (and opinion) in the age of AI' (2024) 53 *Computer Law & Security Review* 105969.

¹⁰² *ibid.* 7

¹⁰³ *ibid.*, p. 8.

examines how AI-driven systems structurally transform the right to privacy under Article 17 ICCPR and Article 8 ECHR. More specifically, it argues that the “inner sphere” protected by privacy rights increasingly encompasses cumulative, relational, and psychologically embedded harms affecting autonomy, behavioral freedom, and identity formation.¹⁰⁴ Accordingly, this thesis does not argue that AI merely threatens cognitive liberty in the abstract, but rather that existing proportionality analysis under international privacy law remains insufficiently equipped to recognize and evaluate these structural forms of interference.

3.1 From Informational Privacy to Behavioral Inference

The deployment of AI-driven systems marks a paradigmatic shift in the nature of privacy interferences. Traditionally, doctrinal protections under Article 17 ICCPR and Article 8 ECHR were primarily designed to shield individuals from the disclosure of sensitive facts, reflecting a “secrecy model” focused upon preventing external access to discrete personal data.¹⁰⁵ However, this section argues that contemporary AI systems transcend this disclosure-centric framework. Rather than merely accessing information intentionally revealed by individuals, AI increasingly operates through behavioral inference, extracting latent characteristics from digital traces never consciously intended for disclosure.¹⁰⁶ As a result, the interference is no longer confined to the “theft” or exposure of personal data alone, but increasingly consists in the unauthorized interpretation, prediction, and reconstruction of the individual’s inner life.¹⁰⁷

3.1.1 Digital Phenotyping

A primary mechanism of this interpretative interference is digital phenotyping, commonly defined as the moment-by-moment quantification of an individual’s human phenotype through smartphone sensors and other digital infrastructures.¹⁰⁸

Lisa Cosgrove and Tom Insel emphasize that these systems focus not upon content itself but rather upon interactional patterns, including typing latency,

¹⁰⁴ *ibid.*

¹⁰⁵ John Tobin and Sarah M Field, ‘Article 16: The Right to Privacy, Family, Home, Correspondence, Honour and Reputation’ in Jonathan Todres and Shani M King (eds), *The Oxford Handbook of Children’s Rights Law* (Oxford University Press 2020) p. 572.

¹⁰⁶ Human Rights Council, *The right to privacy in the digital age: Report of the United Nations High Commissioner for Human Rights* (2021) UN Doc A/HRC/48/31, para 16.

¹⁰⁷ *ibid.*

¹⁰⁸ L Cosgrove and others, ‘Digital Phenotyping and Digital Psychotropic Drugs: Mental Health Surveillance Tools That Threaten Human Rights’ 22(2) *Health and Human Rights Journal*, <<https://www.hhrjournal.org/2020/12/08/perspective-digital-phenotyping-and-digital-psychotropic-drugs-mental-health-surveillance-tools-that-threaten-human-rights/>> accessed 3 May 2026.

scrolling speed, movement patterns, and location data.¹⁰⁹ Their analysis demonstrates how AI systems increasingly derive meaning from behavioral residues rather than from consciously disclosed information.

Similarly, Neil Richards conceptualizes these datasets as a “virtual transcript of the operation of a human mind,” in which ordinary digital behavior functions as a persistent proxy for internal cognition, emotional states, and behavioral tendencies.¹¹⁰ Richards’ argument expands privacy harms beyond informational secrecy and situates them within the broader protection of intellectual and mental self-development.

Building upon these insights, this thesis argues that digital phenotyping increasingly collapses the distinction between the *forum externum* (measurable behavior) and the *forum internum* (mental states).¹¹¹ By transforming involuntary digital traces into “digital biomarkers” for mood, cognition, and behavioral prediction, AI systems perform a form of automated interrogation of the private sphere that bypasses both subjective awareness and meaningful informational self-determination.¹¹² The resulting interference therefore extends beyond external observation and increasingly implicates the conditions under which autonomy and identity are formed.

3.1.2 Predictive Analytics

The inferential capacity of AI systems generates profound asymmetries within the communicative and informational environment.

The Office of the High Commissioner for Human Rights (OHCHR) observes that AI systems permit increasingly far-reaching conclusions concerning highly sensitive matters, including political beliefs, religious affiliation, emotional conditions, and physical or mental vulnerabilities.¹¹³ This illustrates how AI systems transform seemingly neutral digital activity into highly intimate forms of knowledge production.

Building upon this concern, Sue Anne Teo (drawing upon Aswad’s concept of “granular epistemic power”) argues that AI-driven platforms possess unequal forms of knowledge that subsequently generate unequal forms of power.¹¹⁴ Teo’s contribution shifts the analysis away from direct violations alone and toward the structural conditions under which autonomy itself becomes vulnerable to algorithmic influence.

Building on Teo’s framework, I argue that AI systems increasingly create a condition of epistemic asymmetry in which platforms may be placed in a position of epistemic asymmetry, enabling them to infer patterns, vulnerabilities, or behavioral tendencies that individuals themselves may not

¹⁰⁹ *ibid.*, p. 34.

¹¹⁰ Richards, (n 19) p. 389.

¹¹¹ Istace, (n 53) p. 231.

¹¹² Cosgrove, (n 67) p. 34

¹¹³ Human Rights Council, *The right to privacy in the digital age* (2021) UN Doc A/HRC/48/31, para. 16.

¹¹⁴ Teo, (n 2) p. 4

fully recognize or be able to contest.¹¹⁵ In legal terms, this asymmetry directly implicates the “quality of law” requirement underlying international privacy doctrine because the inferential parameters through which individuals are categorized, evaluated, and predicted frequently remain hidden within opaque algorithmic systems.¹¹⁶ As a result, individuals become unable to meaningfully foresee or regulate the conditions under which they are judged.¹¹⁷ This disempowerment produces what may be described as an epistemic crisis, gradually eroding the practical capacity for meaningful human agency and autonomous self-development.¹¹⁸

3.1.3 Emotional Inference

AI systems increasingly utilize emotional recognition technologies premised upon the controversial assumption that internal emotional states may be reliably inferred from external indicators such as facial expressions, vocal patterns, or behavioral cues.¹¹⁹

Sandra Wachter observes that when such inferential practices become embedded within recommender systems, individuals are sorted into “algorithmic groups,” including categories such as “sad teens,” “compulsive consumers,” or “gamblers”. Wachter demonstrates how AI systems increasingly categorize individuals according to predicted vulnerabilities and behavioral tendencies rather than fixed or consciously disclosed identities.¹²⁰

Similarly, the Council of Europe has warned that machine-learning systems increasingly possess the capacity not merely to predict behavior, but also to influence emotions, thoughts, and decision-making processes, often through subliminal or imperceptible mechanisms.¹²¹

Building upon these insights, this thesis argues that algorithmic groupings increasingly function as a form of “artificial immutability.”¹²² Because the criteria governing these classifications — including subtle behavioral indicators, mouse movements, or inferential correlations — frequently remain inaccessible to human understanding, individuals lack meaningful control over the standards through which they are categorized and evaluated.¹²³ Consequently, the private sphere risks being transformed from a space of autonomous self-development into what may be described as a “modulated society,” in which behavior is continuously tuned, conditioned, and optimized at scale for commercial, political, or institutional purposes.¹²⁴

¹¹⁵ Cohen, (n 55) 1905

¹¹⁶ Warso, (n 35), p. 75

¹¹⁷ *ibid.*

¹¹⁸ Y Shany, *The Need for and Feasibility of an International AI Bill of Human Rights* (White Paper, Institute for Ethics in AI, University of Oxford, November 2025)

¹¹⁹ A/HRC/48/31

¹²⁰ Sandra Wachter, ‘The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law’ (2022–2023) 97 *Tulane Law Review* forthcoming

¹²¹ Teo, (n 2) p. 4

¹²² Wachter, (n 120) p. 8

¹²³ *ibid.*, p. 44.

¹²⁴ Cohen, (n 55) pp. 1917.

3.1.4 Boundary Management

Julie Cohen conceptualizes privacy as the “breathing room” necessary for socially situated individuals to engage in processes of boundary management and self-articulation away from external observation.¹²⁵ Cohen frames privacy not merely as secrecy, but as a structural condition necessary for identity formation and autonomous self-development. Building upon this framework, Cohen further argues that the transition toward informationally transforms privacy from a matter of maintaining boundaries into a broader struggle against informational transparency, an attempt by digital systems to eliminate unpredictability and channel human behavior into increasingly measurable and governable trajectories.¹²⁶

I argue that such developments reflect a structural transformation in the nature of privacy harm itself. Traditional interferences were often salient, exogenous, and individually identifiable. AI-driven interferences, by contrast, increasingly operate through cumulative, diffuse, and gradual forms of “slow violence” that subtly reshape behavioral environments, dampen the force of human aspirations, and condition the exercise of autonomy over time.¹²⁷ Accordingly, if AI-driven harms are structurally embedded and psychologically mediated, proportionality analysis under IHRL cannot remain confined to discrete acts of data extraction alone. Rather, it must increasingly scrutinize the broader “conditions of possibility” necessary for autonomy, identity formation, and meaningful human self-development.¹²⁸

3.2 Chilling Effects and Behavioral Modification

A chilling effect is traditionally understood as a form of self censorship, whereby individuals refrain from lawful or desirable activities due to a fear of legal sanction or privacy related harm.¹²⁹ However, beyond this conventional focus on deterrence, scholarship suggests that chilling effects are not merely suppressive, but also productive in that they actively shape behavior.¹³⁰ In contexts of situational uncertainty, such as awareness of surveillance or the presence of vague legal standards, individuals may seek social proof in the conduct of others and adjust their own speech and actions accordingly.¹³¹ In this way, chilling effects operate not only by discouraging expression, but by subtly aligning individual behavior with perceived social norms. Over time, such dynamics risk fostering greater conformity and compliance, with broader implications for moral autonomy, trust, and the quality of democratic deliberation.¹³²

¹²⁵ Cohen, (n 57) p. 5.

¹²⁶ Cohen, (n 55) p. 1911.

¹²⁷ Teo, (n 18) p. 5.

¹²⁸ Teo, (n 57) p. 8.

¹²⁹ Penney, (n 10) p. 1454.

¹³⁰ *ibid.*, p. 1455.

¹³¹ *ibid.*, p. 1488.

¹³² *ibid.*, p. 1458.

In this sense, the impact of AI is not limited to discrete or visible harms but unfolds more gradually through what Sue Anne Teo describes as a form of “slow violence”¹³³, where the cumulative and often imperceptible effects of technological systems begin to reshape the conditions under which human rights are exercised.¹³⁴

Rather than manifesting as a single, identifiable intrusion, these effects operate over time, influencing individuals in subtle ways that may not be immediately apparent but nonetheless carry significant implications for autonomy and self-expression.¹³⁵ As such, the concern is not only whether privacy is directly interfered with, but whether the broader environment in which individuals think, decide, and interact is being continuously shaped in ways that undermine the foundations of human dignity.

Building upon Jonathan Penney’s social conformity theory, the harms generated within AI-mediated environments are often psychological and structural in nature.¹³⁶

Unlike traditional theories centred upon the fear of legal punishment, Penney argues that the mere awareness of being monitored may itself produce acts of social conformity. His theory is particularly significant because it shifts the focus from direct coercion toward behavioural adaptation under conditions of perceived observation.

This dynamic is further reflected in empirical research concerning the “Watching Eye” effect, which demonstrates that even artificial cues of observation may encourage compliant behaviour. Within a “modulated society,” individuals increasingly engage in anticipatory conformity, gradually inclining their choices toward the “bland and the mainstream” in order to avoid negative categorisation under the algorithmic gaze.¹³⁷ Such pressures constitute a profound interference with the “inner personal core” necessary for creativity, experimentation, and rebellion.

3.3 Algorithmic Vulnerability

To understand why this structural harm must be accounted for in law, we must examine the groups most impacted. As proposed by Sue Anne Teo, this refers to a “condition of heightened precarity” in which an individual’s ability to understand, navigate, or challenge algorithmically mediated environments is diminished.¹³⁸

In this context, vulnerability can no longer be treated as a fixed or predefined category. Rather than attaching to particular groups, it becomes dynamic and emergent, shaped by the interaction between individuals and

¹³³ Teo, (n 18)

¹³⁴ Ibid., p. 2265.

¹³⁵ Ibid., p. 2268.

¹³⁶ Penney, (n 10) p. 1451.

¹³⁷ Ibid., p. 1515

¹³⁸ Sue Anne Teo, ‘Artificial intelligence, human vulnerability and multi-level resilience’ (2025) 57 *Computer Law & Security Review* 106134, DOI:<<https://doi-org.ludwig.lub.lu.se/10.1016/j.clsr.2025.106134>> p. 6.

technological systems.¹³⁹ AI does not merely exploit pre existing vulnerabilities; it can also generate new ones through the quantity, intensity, frequency, and quality of exposure to algorithmic stimuli.¹⁴⁰

Within such environments, vulnerability is characterised by a pronounced imbalance of power and knowledge. Individuals are placed in a position of informational disadvantage, while being increasingly represented as a technical set of data to be processed and managed. This shift raises deeper concerns for the protection of autonomy, as the capacity to understand, resist, or meaningfully engage with these systems becomes progressively constrained.¹⁴¹

3.3.1 Children and Youth in the Algorithmic System

Children and youth represent the most significant group bearing the brunt of these risks because they spend a “significant portion of their lives online” and are in a critical stage of developmental exploration.¹⁴²

These concerns are particularly acute in relation to children, whose developmental vulnerability is expressly recognised in the EU AI Act.¹⁴³ The Act prohibits AI systems that exploit a minor’s vulnerability where this is likely to materially distort their behaviour and cause significant harm.¹⁴⁴ This is important because, in the case of children, behavioural distortion cannot be understood only as a temporary change in conduct. Where algorithmic nudging is repeated, personalised, and sustained over time, it may interfere with the child’s developmental trajectory itself, shaping preferences, emotional responses, and patterns of self understanding before these are fully formed.¹⁴⁵ In this sense, the harm engages not only data protection concerns, but the deeper protection of personality, autonomy, and mental integrity.

The paradox of AI companionship illustrates this risk. Generative AI chatbots such as Replika or Character.AI are designed to simulate emotional connection through anthropomorphic and responsive interaction.¹⁴⁶ For young users, this can create a relationship that feels intimate, adaptive, and emotionally validating. Yet the same design logic may also produce dependency or reinforcement effects, particularly where the system learns from a teenager’s speech and responds in ways that prioritise engagement over truth, challenge, or care. In such cases, apparent empathy may become sycophancy, validating delusions, negative core beliefs, or harmful emotional states in order to sustain user attachment.¹⁴⁷ The concern is therefore not

¹³⁹ *ibid.*, p. 5.

¹⁴⁰ *ibid.*, p. 7.

¹⁴¹ Shany, (n 118) p. 63.

¹⁴² United Nations, *Global Principles for Information Integrity: Recommendations for Multi-stakeholder Action* (August 2024), DOI:< <https://doi.org/10.18356/9789211065756>> accessed 1 May 2026, p. 9.

¹⁴³ *Article 5: Prohibited AI Practices* (EU Artificial Intelligence Act)

¹⁴⁴ Shany, (n 118) p. 25.

¹⁴⁵ Teo (n 18) p. 5.

¹⁴⁶ Teo, (n 138) p. 8.

¹⁴⁷ *ibid.*, p. 8.

merely that the child uses AI, but that the system may become embedded in the child's emotional world in ways that distort judgment, dependency, and mental integrity.¹⁴⁸

This also affects identity formation. Childhood and adolescence require a private sphere in which individuals can experiment, make mistakes, change their minds, and develop a sense of self without every action becoming a permanent data point. However, where children know that they are being ranked, profiled, assessed, or measured by educational algorithms, social media systems, or other data driven environments, the chilling effect becomes especially severe.¹⁴⁹ They may avoid trial and error, suppress uncertain thoughts, or conform to perceived expectations because they fear that present behaviour will follow them into adulthood.¹⁵⁰ This produces a collapse of identity exploration: the child's future self is shaped not through free development, but through anticipatory self discipline under conditions of continuous algorithmic observation.

3.3.2 Adults in Crisis

These concerns are not limited to children. Adults experiencing conditions such as mental health crises, economic precarity, or diminished impulse control may be equally exposed to structural forms of privacy harm.¹⁵¹ In such contexts, vulnerability arises not as a fixed status, but as a situational condition that can be identified, amplified, and exploited through AI driven systems.

One manifestation of this is the targeting of individuals with low impulse control.¹⁵² As Sue Anne Teo illustrated in her paper, machine learning models are capable of analyzing behavioral indicators, including language patterns, gestures, and eye movements, in order to infer susceptibility.¹⁵³ These inferred characteristics may then be leveraged through subliminal techniques or highly personalized forms of advertising that steer individuals towards decisions that are contrary to their best interests, such as compulsive consumption or gambling.¹⁵⁴ In this way, the interference extends beyond access to information and engages the individual's capacity for self control, decision making, and mental integrity.

¹⁴⁸ I Steenstra and others, 'Assessing Risks of Large Language Models in Mental Health Support: A Framework for Automated Clinical AI Red Teaming' (2026) DOI:<<https://doi.org/10.48550/arXiv.2602.19948>> accessed 3 May 2026.

¹⁴⁹ S A Teo, 'Human dignity and AI: mapping the contours and utility of human dignity in addressing challenges presented by AI' (2023) 15(1) *Law, Innovation and Technology*, DOI:< <https://doi.org/10.1080/17579961.2023.2184132>> p. 257.

¹⁵⁰ Sue Anne Teo, Artificial intelligence and its 'slow violence' to human rights. *AI Ethics* 5, 2265–2280 (2025). DOI: <<https://doi.org/10.1007/s43681-024-00547-x>> p. 2270.

¹⁵¹ UN Human Rights Council, *The Right to Privacy in the Digital Age* (Report of the United Nations High Commissioner for Human Rights, 48th sess, 13 September–1 October 2021) UN Doc A/HRC/48/31, para. 4.

¹⁵² Teo, (n 138) p. 7.

¹⁵³ *ibid.*

¹⁵⁴ Shany, (n 118) p. 13.

A related concern arises in the context of public administration and welfare systems.¹⁵⁵ AI is increasingly used for purposes of risk evaluation and referral triage, often relying on extensive data matching across multiple domains.¹⁵⁶ Such practices may generate an unprecedented level of intrusiveness into the lives of individuals who are already socially and economically vulnerable.¹⁵⁷ These concerns were illustrated in the Netherlands, where a court ruling banned a digital welfare fraud detection system for infringing the right to privacy.¹⁵⁸ The system enabled extensive data-sharing between central and local authorities, combining information on employment, housing, education, social benefits and health insurance, and was deployed primarily in low-income and minority neighborhoods, resulting in de facto socioeconomic discrimination. In this way, such systems reduce individuals to technical representations within data infrastructures, obscuring their status as moral agents and contributing to structural harms that unfold gradually over time.¹⁵⁹

Finally, significant risks emerge in the use of AI within mental health support. Individuals who turn to AI based therapeutic systems may be exposed to forms of care that lack the intersubjective grounding of human relationships.¹⁶⁰ In particular, concerns have been raised that such systems may fail to appropriately respond to situations of acute distress, including suicidal ideation.¹⁶¹ The replacement of human interaction with simulated forms of empathy may not only limit the effectiveness of care, but also risk reinforcing isolation and dependency.¹⁶² In this context, the deployment of AI does not merely raise questions of technological reliability, but engages deeper issues relating to autonomy, dignity, and the conditions necessary for meaningful human support.

In conclusion, traditional proportionality analysis struggles to adequately recognize these forms of harm because existing legal frameworks remain primarily oriented toward visible, measurable, and immediate interferences with privacy. When courts treat psychological or structural “chilling effects” as overly speculative for legal recognition, a significant “proportionality gap” emerges, one that risks overlooking the most dangerous features of AI systems: harms that accumulate gradually and only become apparent over time. The following chapter will therefore examine how existing international proportionality doctrine has been applied within cases involving surveillance, algorithmic governance, and technologically mediated interferences with private life, with particular attention to the extent to which current jurisprudence remains capable of recognizing structural harms.

¹⁵⁵ UN Human Rights Council, UN Doc A/HRC/48/31 (n 36) para. 31.

¹⁵⁶ H Molbæk-Steensig and M Scheinin, ‘Human Rights and Artificial Intelligence in Healthcare-Related Settings’ (2025) 32(2) *European Journal of Health Law*, p. 9.

¹⁵⁷ UN Human Rights Council, UN Doc A/HRC/48/31 (n 36) para. 31.

¹⁵⁸ *ibid.*

¹⁵⁹ *ibid.*

¹⁶⁰ Steenstra et al, (n 148)

¹⁶¹ *Ibid.*

¹⁶² Cosgrove et al, (n 67) p. 37.

4 State Responsibility in AI Governance

Building upon the previous analysis of how AI systems enact a form of “slow violence” through behavioral inference and environmental modulation¹⁶³, this chapter examines whether existing proportionality doctrine under Article 17 ICCPR and Article 8 ECHR is capable of recognizing cumulative, structural, and psychologically embedded harms. The central argument advanced is that, although IHRL already provides doctrinal mechanisms capable of addressing such harms particularly through the principles of legality, necessity, and proportionality, there nevertheless exists a profound “proportionality gap.”¹⁶⁴ This gap does not arise from the absence of legal principles themselves, but rather from a jurisprudence that remains largely tethered to a “salient harm” model, one primarily oriented toward visible, temporally proximate, and individually identifiable interferences with privacy.¹⁶⁵ Consequently, diffuse, relational, and structurally embedded erosions of the *forum internum* frequently remain insufficiently articulated within existing proportionality analysis.¹⁶⁶

4.1 Negative Obligation: *The Dutch Systeem Risicoindicatie Case*

The Dutch Systeem Risicoindicatie (SyRI) case serves as a critical paradigm for evaluating state-led algorithmic governance.¹⁶⁷ SyRI was introduced as a statutory mechanism designed to combat fraud in social security and taxation through the large-scale linkage and analysis of datasets obtained from multiple government authorities. While the system pursued a legitimate governmental objective, the case exposed the extent to which algorithmic governance may destabilize the principles of legal certainty, foreseeability, and individual autonomy underlying Article 8 ECHR.¹⁶⁸

The Hague District Court evaluated the SyRI framework against the three-part test established under Article 8(2) ECHR¹⁶⁹, namely:

1. whether the interference was “in accordance with the law”;

Both the State and the court accepted that SyRI’s data exchange and risk analysis constituted an interference with private life;¹⁷⁰

2. whether it pursued a legitimate aim;

¹⁶³ Teo, (n 2) p. 53

¹⁶⁴ *ibid.*

¹⁶⁵ Sue Anne Teo, ‘How Artificial Intelligence Systems Challenge the Conceptual Foundations of the Human Rights Legal Framework’ (2022) 40(1) *Nordic Journal of Human Rights*, p. 216.

¹⁶⁶ Teo, (n 2)

¹⁶⁷ *ibid.*, p. 8.

¹⁶⁸ *NJCM c.s. and others v De Staat der Nederlanden* (District Court of The Hague, 5 February 2020) ECLI:NL:RBDHA:2020:1878 (SyRI). Para. 6.7.

¹⁶⁹ *ibid.*, para. 6.73.

¹⁷⁰ *ibid.*, para. 6.66.

The court ruled that when a State employs new, sophisticated technologies that can interfere extensively with private life, it bears a “special responsibility” to strike a balance that protects against abuse and arbitrariness¹⁷¹

3. whether it was “necessary in a democratic society.”

The court accepted that combating fraud serves the “economic wellbeing of the country” (a legitimate aim) and that there is a “pressing social need” to address social security fraud. However, it concluded that the specific *means* chosen (the SyRI legislation) were not proportionate to that aim.

To analyze in detail, both the State and the Court accepted that SyRI’s extensive data exchange and algorithmic risk analysis constituted an interference with the right to private life under Article 8 ECHR.¹⁷² The central legal issue therefore did not concern the existence of an interference itself, but whether the nature and operation of the SyRI framework satisfied the proportionality requirements necessary to justify such interference.¹⁷³

A particularly significant aspect of the judgment concerns the Court’s articulation of the State’s “special responsibility” when deploying novel and sophisticated technologies capable of extensively interfering with private life.¹⁷⁴ The Court emphasized that where technologically advanced systems generate heightened risks of abuse, arbitrariness, or disproportionate intrusion, States bear an enhanced obligation to ensure that sufficient safeguards exist to protect individuals against unjustified interferences. This reasoning is doctrinally important because it recognizes that algorithmic governance may intensify existing asymmetries of power and information between the State and the individual, thereby requiring stricter scrutiny under Article 8 ECHR.¹⁷⁵

The Court accepted that combating fraud within the social security system pursued the legitimate aim of protecting the “economic wellbeing of the country” and responded to a recognized “pressing social need”.¹⁷⁶ However, the existence of a legitimate objective did not itself render the SyRI framework proportionate. Rather, the Court emphasized that the proportionality analysis required an assessment of whether the specific means employed under the legislation struck a “fair balance” between the public interest in fraud prevention and the protection of private life.¹⁷⁷ Ultimately, the Court concluded that the SyRI legislation failed to satisfy this requirement because the framework lacked adequate transparency, safeguards, and verifiability to prevent arbitrary interference.¹⁷⁸

The significance of this doctrinal baseline lies in the Court’s recognition that proportionality analysis under Article 8 ECHR extends beyond formal

¹⁷¹ *ibid.*, paras. 6.89-6.90.

¹⁷² *ibid.*, para. 6.7.

¹⁷³ *ibid.*

¹⁷⁴ *ibid.*

¹⁷⁵ *ibid.*, para. 6.51.

¹⁷⁶ *ibid.*, paras. 6.74-6.75.

¹⁷⁷ *ibid.*, para. 6.7.

¹⁷⁸ *ibid.*, paras. 6.89-6.90.

statutory authorization alone. The judgment demonstrates that technologically mediated systems of governance require particularly rigorous scrutiny where their operation remains opaque or difficult to challenge.¹⁷⁹ In this sense, SyRI illustrates how algorithmic governance may undermine legal certainty, foreseeability, and protection against arbitrariness. The principles that remain central to proportionality doctrine within IHL.

4.1.1 The Court's Reasoning

A. Foreseeability

The Court's most significant objection concerned the lack of transparency and verifiability within the SyRI framework. Central to this concern was the government's refusal to disclose the risk models, indicators, and weighting mechanisms used to generate risk reports, including the so-called "simple decision tree."¹⁸⁰ The State argued that disclosure would enable individuals to "game the system" and undermine the effectiveness of fraud detection measures. However, the Court concluded that such secrecy fundamentally weakened the foreseeability requirement underlying Article 8 ECHR.¹⁸¹

In particular, the Court found that the opaque nature of the SyRI system made it impossible for individuals to understand why they had been flagged as presenting a heightened fraud risk or to meaningfully defend themselves against a risk report.¹⁸² This reasoning is doctrinally significant because foreseeability under Article 8 ECHR requires more than mere statutory authorization; individuals must also possess a sufficient understanding of how legal powers are exercised in practice. Where algorithmic decision-making processes remain inaccessible, the individual becomes unable to regulate their conduct in response to the standards governing state suspicion.¹⁸³

This lack of transparency also generated a pronounced form of epistemic asymmetry between the State and the data subject.¹⁸⁴ Because individuals were neither notified that their personal data had been processed nor informed that a risk report had been generated, they lacked the ability to track, verify, or challenge the inferential conclusions produced about them.¹⁸⁵ Consequently, the Court recognized that the SyRI framework undermined informational self-determination by depriving individuals of meaningful control over the processing and interpretation of their personal data.¹⁸⁶

¹⁷⁹ *ibid.*, para. 6.7.

¹⁸⁰ *ibid.*, para. 6.47.

¹⁸¹ *ibid.*, para. 6.72.

¹⁸² *ibid.*, paras. 6.89, 6.90, 6.94. The Court specifically notes that the lack of insight into the validation of risk models and indicators means it is "difficult to comprehend how a data subject could be able to defend themselves".

¹⁸³ *ibid.*, paras. 6.66, 6.90, 6.95.

¹⁸⁴ *ibid.*, 6.90.

¹⁸⁵ *ibid.*, paras. 6.54, 6.82, 6.90

¹⁸⁶ *ibid.*, paras. 6.82, 6.90.

The significance of this finding extends beyond procedural opacity alone. The Court’s reasoning demonstrates that algorithmic governance may interfere with legal certainty where individuals cannot meaningfully foresee how ordinary behavioral traces become transformed into indicators of suspicion within a “black box” system.¹⁸⁷ Thus, the interference operates not merely through data collection itself, but through the opaque inferential processes used to construct risk profiles capable of generating significant legal and social consequences.¹⁸⁸

B. Structural Vulnerability

The Court further identified a “structural vulnerability” in the manner in which the SyRI system had been deployed.¹⁸⁹ Although the legislation itself was formally framed as a neutral anti-fraud mechanism, the practical operation of the system revealed a pattern of targeted deployment within low-income or so-called “problem” neighborhoods.¹⁹⁰

This selective deployment raised significant concerns regarding bias, stigmatization, and discriminatory risk.¹⁹¹ The Court observed that the large-scale analysis of personal data within socioeconomically vulnerable areas risked generating “algorithmic groups” structured around statistical assumptions linked to poverty, migration background, or social marginalization.¹⁹² Rather than identifying objective indicators of fraud, the system risked reinforcing pre-existing stereotypes and reproducing patterns of structural disadvantage through automated risk categorization.¹⁹³

The Court’s reasoning demonstrates that proportionality analysis under Article 8 ECHR is capable of recognizing structural vulnerabilities associated with algorithmic governance.¹⁹⁴ At the same time, however, the judgment remained primarily focused upon discriminatory deployment and unequal exposure to surveillance rather than the cumulative psychological effects associated with persistent algorithmic suspicion.¹⁹⁵ Consequently, while the Court recognized that algorithmic systems may intensify existing social inequalities, the broader implications for behavioral autonomy and anticipatory conformity remained comparatively underdeveloped.¹⁹⁶

C. Failures in Data Minimization and Purpose Limitation

The Court additionally found that the SyRI legislation paid insufficient attention to the principles of data minimization and purpose limitation, both

¹⁸⁷ *ibid.*, paras. 6.85, 6.90, 6.95

¹⁸⁸ *ibid.*, paras. 6.62, 6.63, 6.95

¹⁸⁹ *ibid.*, para. 6.93.

¹⁹⁰ *ibid.*, paras. 6.91, 6.93.

¹⁹¹ *ibid.*, paras. 6.92, 6.93.

¹⁹² *ibid.*, para. 6.94.

¹⁹³ *ibid.*, para. 6.92

¹⁹⁴ *ibid.*, paras. 6.84, 6.93-6.95.

¹⁹⁵ *ibid.*, paras. 6.86, 6.95.

¹⁹⁶ *ibid.*, paras. 6.86-6.107.

of which remain central safeguards within European data protection and privacy law.¹⁹⁷

Although the legislation formally identified the categories of personal data capable of being processed under the SyRI framework, the scope of these categories was exceptionally broad. In total, seventeen categories of data were eligible for analysis.¹⁹⁸ Including information relating to work, education, housing, health insurance, and financial circumstances, The Court observed that the breadth of these categories made it “hard to imagine any type of personal data that is not eligible for processing in SyRI.”¹⁹⁹

This overbreadth significantly weakened the proportionality assessment because the legislation failed to sufficiently restrict the scope of interference to what was strictly necessary for the stated anti-fraud objective.²⁰⁰ In particular, the Court noted the absence of meaningful limitations capable of ensuring that only relevant and necessary data would be processed in specific projects.²⁰¹

The Court also attached considerable importance to the absence of independent review mechanisms prior to data processing.²⁰² The legislation did not require a comprehensive assessment by an independent third party capable of determining whether the proposed interference was strictly necessary and proportionate within a particular context. As a result, the system lacked an important procedural safeguard against excessive or arbitrary data processing.²⁰³

The significance of this deficiency lies in the Court’s recognition that proportionality analysis requires more than broad legislative authorization for technological experimentation. Rather, States deploying algorithmic systems must demonstrate why extensive forms of data linkage and predictive analysis are strictly necessary in relation to the legitimate objective pursued.²⁰⁴

D. Insufficiency of Human Oversight

The State argued that SyRI did not constitute a fully automated system because the Social Affairs and Employment Inspectorate conducted a subsequent human assessment of generated risk reports. According to the government, this human intervention functioned as a safeguard capable of preventing arbitrary or biased outcomes.²⁰⁵

The Court rejected this argument. It concluded that because the internal logic underlying the “black box” selection process remained non-public, the subsequent human review could not be meaningfully verified for either accuracy or bias. In other words, the opacity of the underlying algorithmic

¹⁹⁷ *ibid.*, paras. 6.86, 6.96

¹⁹⁸ *ibid.*, paras. 4.17, 6.50, 6.98.

¹⁹⁹ *ibid.*, para. 6.98.

²⁰⁰ *ibid.*, paras. 6.95, 6.106.

²⁰¹ *ibid.*, paras. 6.98, 6.99.

²⁰² *ibid.*, paras. 6.97, 6.99, 6.106.

²⁰³ *ibid.*, paras. 6.99, 6.106.

²⁰⁴ *ibid.*, paras. 6.84, 6.85, 6.102.

²⁰⁵ *ibid.*, paras. 6.58, 6.94.

system fundamentally limited the effectiveness of any later human assessment.²⁰⁶

This reasoning is particularly significant because it challenges the assumption that the mere presence of human involvement automatically resolves concerns regarding algorithmic arbitrariness.²⁰⁷ The judgment suggests that human oversight cannot function as an effective safeguard where the inferential logic guiding algorithmic selection remains inaccessible or insufficiently transparent.²⁰⁸ Consequently, the Court recognized that formal human review mechanisms may become procedurally superficial where decision-makers themselves remain unable to meaningfully interrogate or verify the operation of the underlying system.²⁰⁹

4.2 Positive Obligation: The *Bărbulescu* Proportionality Test

While SyRI addresses direct state interference, *Bărbulescu v. Romania*²¹⁰ serves as the definitive template for the State's positive obligation to secure respect for private life within the structural power imbalances of the private sector.²¹¹ The case concerned the dismissal of an employee following the employer's monitoring of his private Yahoo Messenger communications.²¹² Because the monitoring was conducted by a private company, the ECtHR assessed whether the Romanian State had provided an adequate legal framework and effective judicial remedies capable of protecting the applicant's Article 8 rights.²¹³

The Grand Chamber found that the Romanian domestic courts had adopted an excessively "rigid approach" by largely deferring to the employer's decision without conducting an independent and sufficiently rigorous proportionality assessment under Article 8 ECHR.²¹⁴ In particular, the domestic courts failed to adequately balance the applicant's right to private life against the employer's interest in maintaining workplace discipline and productivity.²¹⁵ To remedy this deficiency, the Grand Chamber articulated a detailed six-factor proportionality framework designed to assess whether the State had fulfilled its positive obligation to secure a fair balance between competing interests within the employment context.²¹⁶

First, the Court emphasized the importance of prior notification. Domestic authorities were required to determine whether the employee had received

²⁰⁶ *ibid.*, para. 6.94

²⁰⁷ *ibid.*, paras. 6.90, 6.94.

²⁰⁸ *ibid.*, paras. 6.94, 6.95

²⁰⁹ *ibid.*, para. 6.94.

²¹⁰ *Bărbulescu v Romania* App no 61496/08 (ECtHR, Grand Chamber, 5 September 2017)

²¹¹ *ibid.*, paras. 108-111.

²¹² *ibid.*, paras. 11, 23, 109.

²¹³ *ibid.*, paras. 110, 115, 116.

²¹⁴ *ibid.*, paras. 89, 139, 141.

²¹⁵ *ibid.*, paras. 140, 141.

²¹⁶ *ibid.*, para. 121.

clear and advance notice concerning both the possibility and the nature of workplace monitoring.²¹⁷ This requirement was considered essential because individuals cannot reasonably regulate their conduct where surveillance measures remain ambiguous or undisclosed.²¹⁸

Second, the Court stressed that a fundamental distinction must be drawn between monitoring the flow of communications, such as metadata, and accessing the actual content of communications. The judgment recognized that content monitoring constitutes a significantly more intrusive interference with private life because it permits direct access to the substance of an individual's personal relationships, thoughts, and private exchanges.²¹⁹

Third, the Court held that employers must provide "weightier justifications" for particularly invasive forms of monitoring, especially where the content of private communications is accessed. The proportionality analysis therefore required more than a generalized interest in workplace efficiency; rather, the employer was expected to demonstrate a sufficiently serious and specific justification capable of warranting the level of intrusion imposed.²²⁰

Fourth, the Grand Chamber emphasized the principle of least intrusive means.²²¹ Domestic authorities were required to assess whether the employer's legitimate objectives, such as ensuring productivity or protecting company resources, could have been achieved through measures less intrusive than direct access to the applicant's personal messages. This aspect of the judgment is particularly significant because it demonstrates that proportionality analysis under Article 8 ECHR requires an evaluation not only of legitimacy, but also of necessity and the availability of less restrictive alternatives.²²²

Fifth, the Court stressed that the consequences of the monitoring for the employee must form part of the balancing exercise. In *Bărbulescu*, the applicant's dismissal represented the most severe disciplinary sanction available, thereby intensifying the seriousness of the interference with his private life and professional existence.²²³

Finally, the Grand Chamber examined whether the monitoring was accompanied by adequate and sufficient safeguards against abuse.²²⁴ Such safeguards included limitations on access to content, procedural protections surrounding the use of surveillance measures, and mechanisms ensuring that intrusive monitoring would not occur arbitrarily or disproportionately. The Court specifically noted the importance of ensuring that access to the content

²¹⁷ *ibid.*, para. 121.

²¹⁸ *ibid.*, paras. 121, 133

²¹⁹ *ibid.*, paras. 72, 121.

²²⁰ *ibid.*, para. 135.

²²¹ *ibid.*, para. 121(iv).

²²² *ibid.*, paras. 136, 140.

²²³ *ibid.*, para. 137.

²²⁴ *ibid.*, para. 121(vi).

of communications should ordinarily occur only after the employee had been informed and provided an opportunity to respond.²²⁵

Applying this framework, the Grand Chamber concluded that the Romanian courts had failed to determine at what stage the employer had accessed the content of the applicant's communications and whether less intrusive alternatives had been available.²²⁶ Consequently, the domestic authorities failed to provide sufficient protection against arbitrary interference with the applicant's Article 8 rights.²²⁷ The judgment is therefore doctrinally significant because it demonstrates that proportionality analysis under Article 8 ECHR extends beyond formal employer policies and requires a substantive assessment of necessity, intrusiveness, and safeguards within technologically mediated monitoring environments.²²⁸

4.3 Synthesis: Structural Harms and the Proportionality Gap

The preceding analysis of *SyRI* and *Bărbulescu v Romania* reveals a fundamental paradox within contemporary human rights adjudication. While IHRL possesses the doctrinal sophistication necessary to identify procedural irregularities within algorithmic systems, its application remains insufficiently equipped to account for the structural transformation of the informational environment itself. The "proportionality gap" identified throughout this chapter does not arise from the absence of legal principles, but rather from a form of jurisprudential under-inclusivity that continues to evaluate AI-mediated interferences primarily through a "salient harm" model.²²⁹ Consequently, proportionality analysis remains largely oriented toward visible, temporally proximate, and individually identifiable interferences, while diffuse, cumulative, and structurally embedded erosions of the *forum internum* frequently remain insufficiently articulated within existing jurisprudence.

The preceding case studies demonstrate that existing proportionality doctrine is already capable of recognizing concerns relating to opacity, arbitrariness, safeguards, and informational asymmetries. However, both *SyRI* and *Bărbulescu* ultimately remained tethered to procedural and externally visible forms of interference. In *SyRI*, the Hague District Court focused primarily upon the opacity of the "black box" system and the overbreadth of data processing as procedural deficiencies.²³⁰ Similarly, in *Bărbulescu*, the Grand Chamber centered much of its reasoning upon the absence of prior notice and procedural safeguards surrounding workplace

²²⁵ *ibid.*, para. 138.

²²⁶ *ibid.*, para. 140.

²²⁷ *ibid.*, para. 141.

²²⁸ *ibid.*, para. 141;

²²⁹ Teo, (n 165) p. 218

²³⁰ *NJCM et al. v. The Netherlands (SyRI)*, (n168) paras. 6.90, 6.98.

monitoring.²³¹ While doctrinally significant, these approaches continue to conceptualize privacy harms largely as discrete “events” rather than as processes of environmental modulation capable of gradually reshaping the conditions under which autonomy and self-development occur.²³²

Building on Julie Cohen’s conception of privacy as the “breathing room” necessary for socially situated subjects to engage in boundary management and self-articulation away from the omnipresent gaze²³³, this thesis argues that proportionality analysis must move beyond a narrow “violations vernacular” focused solely on identifiable intrusions. In AI-mediated environments, the interference often lies not in a single act of disclosure or surveillance, but in the cumulative psychological pressure generated through persistent inferential monitoring, anticipatory conformity, and behavioral modulation.²³⁴ Accordingly, safeguarding the “inner sphere” under Article 8 ECHR and Article 17 ICCPR requires a model of structural scrutiny capable of evaluating not merely isolated interferences, but also the broader “conditions of possibility” for autonomous self-development.²³⁵

Building upon the preceding discussion, bridging the proportionality gap requires more than the formal existence of safeguards within AI governance frameworks.²³⁶ Although existing proportionality doctrine under Article 8 ECHR and Article 17 ICCPR already contains mechanisms capable of scrutinizing opacity, arbitrariness, and disproportionate interferences, the analysis often remains confined to procedural legality rather than the substantive conditions under which autonomy is exercised.²³⁷ The Court’s invocation of the State’s “special responsibility” in *SyRI* is particularly significant in this regard because it implicitly recognizes that technologically sophisticated systems capable of extensive interference with private life demand a heightened level of scrutiny beyond ordinary administrative decision making.²³⁸

This heightened scrutiny becomes necessary because AI mediated interferences do not arise solely through the disclosure or collection of personal information.²³⁹ Rather, the interference increasingly operates through inferential processes that transform ordinary behavioral traces into predictive assessments, vulnerability classifications, and risk profiles capable of shaping how individuals are governed, evaluated, and perceived.²⁴⁰ Where such inferential processes remain opaque, individuals are deprived not only of informational control, but also of the practical ability to understand how suspicion, categorization, or behavioral evaluation is constructed.²⁴¹

²³¹ *Bărbulescu v. Romania*, (n 210) paras. 121, 141.

²³² *Teo*, (n 165) p. 216.

²³³ *Cohen*, (n 55) p. 1911.

²³⁴ *Penney*, (n 10) p. 1491.

²³⁵ *Teo*, (n 165) pp. 216, 218.

²³⁶ *ibid.*, p. 216.

²³⁷ *ibid.*

²³⁸ *NJCM et al. v. The Netherlands (SyRI)*, (n 168) paras. 6.84, 6.85.

²³⁹ *Teo*, (n 165) p. 218.

²⁴⁰ *Teo*, (n 2) p. 5; 30, *Wachter*, (n 102) p. 44.

²⁴¹ *Shany*, (n 118) p. 8.

Consequently, the interference affects the principle of legal certainty itself, since foreseeability under Article 8 ECHR requires more than the mere existence of statutory authorization; it also requires that individuals possess a meaningful ability to comprehend how legal powers are exercised in practice.²⁴²

The limitations of current proportionality analysis therefore become particularly visible where safeguards remain predominantly procedural in nature.²⁴³ Transparency obligations, notification requirements, and forms of human oversight undoubtedly remain doctrinally important. However, as both *SyRI* and *Bărbulescu* demonstrate, such safeguards may become insufficient where the underlying architecture of algorithmic governance continues to operate through persistent epistemic asymmetry and behavioral modulation.²⁴⁴ In such contexts, the existence of formal notice does not necessarily restore meaningful autonomy where individuals remain unable to challenge the inferential logic shaping their digital environment or escape forms of continuous algorithmic evaluation.²⁴⁵

Accordingly, the “fair balance” required under proportionality analysis must extend beyond the balancing of administrative efficiency against informational privacy interests alone.²⁴⁶ It must also account for the cumulative psychological and structural effects produced through persistent systems of algorithmic observation and prediction.²⁴⁷ Otherwise, proportionality review risks remaining tethered to a “salient harm” model that recognizes only visible and temporally proximate injuries, while diffuse erosions of decisional autonomy, behavioral freedom, and self-development remain insufficiently articulated within existing jurisprudence.²⁴⁸

Viewed in this light, structural scrutiny does not require abandoning existing proportionality doctrine, but rather applying its established principles of legality, necessity, safeguards, and fair balance in a manner capable of recognizing cumulative and psychologically embedded harms as legally relevant interferences with private life.²⁴⁹ Only through such an approach can Article 8 ECHR and Article 17 ICCPR remain capable of safeguarding the integrity of the *forum internum* within increasingly pervasive systems of algorithmic governance.²⁵⁰

4.4 AI Governance and the Limits of Current Regulatory Frameworks

²⁴² Warso, (n 35) p. 75.

²⁴³ *ibid.*, p. 79.

²⁴⁴ *ibid.*, p. 75; *Syri* para. 6.94.

²⁴⁵ Teo, (n 2) p. 6; Wachter, (n 120) pp. 43-44.

²⁴⁶ Teo, (n 165) p. 216.

²⁴⁷ Penney, (n 10) pp. 1455, 1491.

²⁴⁸ Teo, (n 165) pp. 216, 218.

²⁴⁹ *ibid.*, p. 218.

²⁵⁰ Istace, (n 53) p. 214.

The preceding analysis of the *SyRI* and *Bărbulescu* cases demonstrates a critical tension: while IHRL possesses the doctrinal machinery to scrutinize algorithmic systems, its application often fails to account for the structural and internal transformation of the informational environment. This section examines the current shift within States and international institutions from voluntary AI ethics toward binding human-rights-based regulation. I argue that while these emerging frameworks increasingly recognize the unique risks of AI, such as opacity, behavioral manipulation, and epistemic asymmetry, they remain primarily anchored in a model of procedural compliance that is under-inclusive of the diffuse and psychologically embedded harms affecting the “inner sphere”.

4.4.1 The Transition to Binding Human-Rights-Based Regulation

International institutions and States are increasingly moving beyond voluntary “AI Ethics” principles toward binding legislative frameworks grounded in IHRL.²⁵¹ The EU AI Act²⁵² represents a significant shift within this transition by transforming previously aspirational principles into enforceable legal obligations through a risk based regulatory model that classifies AI systems according to their potential impact upon health, safety, and fundamental rights. In parallel, the Council of Europe Framework Convention on AI further reflects this movement toward rights based governance by requiring Parties to ensure that AI systems throughout their lifecycle remain compatible with human rights, democracy, and the rule of law.²⁵³

This broader regulatory transition is reinforced by an expanding body of IHRL guidance recognizing that AI systems intensify privacy interferences in ways that exceed traditional models of data collection and disclosure.²⁵⁴ OHCHR reports on privacy in the digital age emphasize that AI driven technologies deepen risks associated with surveillance, profiling, and behavioral inference, thereby reaffirming the continuing importance of legality, necessity, and proportionality within contemporary systems of governance.²⁵⁵ Similarly, recent reports from the UN Working Group on Business and Human Rights clarify that the UNGPs require both States and

²⁵¹ UN Working Group on Business and Human Rights, *The Practical Application of the Guiding Principles on Business and Human Rights to the Activities of Technology Companies* UN Doc A/HRC/59/53 (2025) paras 2, 46; Helga Molbæk-Steensig and Martin Scheinin, ‘Human Rights and Artificial Intelligence in Healthcare-related Settings’ in Tine Hanrieder, Aadya Sharma and Paul B Martins (eds), *The Future of Human Rights Governance* (Cambridge University Press 2023), p. 3.

²⁵² *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)* [2024] OJ L 2024/1689.

²⁵³ Molbæk-Steensig and Scheinin, (n 156), p. 4.

²⁵⁴ Office of the United Nations High Commissioner for Human Rights, *The Right to Privacy in the Digital Age* UN Doc A/HRC/48/31 (2021) para 12;

²⁵⁵ *ibid.*, para. 38.

private enterprises to integrate Human Rights Due Diligence (HRDD) throughout the AI lifecycle in order to identify, prevent, and mitigate harms associated with exploitative surveillance and the manipulation of consumer behavior.²⁵⁶ Consequently, emerging regulatory frameworks increasingly rely upon mechanisms such as transparency obligations, human oversight requirements, mandatory Fundamental Rights Impact Assessments (FRIAs), and algorithm registers as safeguards intended to constrain the risks generated through AI mediated systems.²⁵⁷

4.4.2 The “Salient Harm” Model

Despite these developments, existing regulatory and proportionality frameworks remain insufficiently equipped to articulate the cumulative, structural, and psychologically embedded harms of the AI era.²⁵⁸ Most contemporary governance strategies, including the risk management systems mandated by the EU AI Act, remain oriented toward procedural compliance and the mitigation of “reasonably foreseeable” risks. While such mechanisms are important, their procedural focus risks reducing fundamental rights protection to a “checkbox” exercise, where transparency and explainability are treated as ends in themselves rather than as tools for preserving the conditions of possibility for autonomy.²⁵⁹

This limitation reflects the deeper problem identified by Sue Anne Teo: IHRL remains ill-suited to AI harms because its harm typology is misaligned.²⁶⁰ Traditional human rights protection is generally structured around “salient harms”, namely harms that are exogenous, observable, and materially visible.²⁶¹ AI harms, by contrast, are often internal, hidden, and latent.²⁶² For instance, while the EU AI Act prohibits “subliminal techniques” that cause “significant harm,” the threshold of “significance” is still likely to be interpreted through a salient-harm lens. Such an approach may fail to capture the slow violence of gradual behavioral conditioning and anticipatory conformity.²⁶³ By focusing primarily on immediate material or physical injuries, regulatory frameworks risk overlooking the structural attrition of the private sphere, where the harm is not a discrete violation but the persistent herding of behavior toward a normalized social core.²⁶⁴

²⁵⁶ UN Working Group on Business and Human Rights, *The Practical Application of the Guiding Principles on Business and Human Rights to the Activities of Technology Companies* UN Doc A/HRC/59/53 (2025) paras. 2,4.

²⁵⁷ A/HRC/48/31, (n 236) para. 55.

²⁵⁸ Teo, (n 165) pp. 216, 218.

²⁵⁹ Teo, (n 2) p. 63.

²⁶⁰ Teo, (n 165) p. 216.

²⁶¹ *ibid.*, p. 218.

²⁶² *ibid.*, p. 218; Ian Steenstra, Paola Pedrelli, Weiyan Shi, Stacy Marsella and Timothy W Bickmore, ‘Assessing Risks of Large Language Models in Mental Health Support: A Framework for Automated Clinical AI Red Teaming’ (2026)

²⁶³ Penney, (n 10) p. 1491.

²⁶⁴ Teo, (n 165) p. 216.

4.4.3 The Resilience of the Inner Sphere

This under-inclusivity is further intensified by the epistemic asymmetry inherent in “black box” algorithmic governance.²⁶⁵ Even where transparency is formally mandated, “transparency without control” may remain meaningless for the individual.²⁶⁶ As Julie Cohen notes, privacy is the opposite of modulation; it requires “semantic discontinuity” and the “breathing room” necessary for boundary management away from an omniscient gaze.²⁶⁷

Current regulatory efforts often fail to protect the “inner sphere” because they do not substantively restrict the granular epistemic power that platforms and States wield to identify and exploit individual vulnerabilities.²⁶⁸ Instead of questioning the “logic of data extraction” itself, current proportionality reviews often settle for notice-and-consent models, which effectively allow the structural conditions of autonomy to be traded away for access to essential services.²⁶⁹ The core limitation, therefore, is not the absence of doctrinal mechanisms, but a jurisprudential under-inclusivity that fails to recognize diffuse psychological attrition and environmental modulation as cognizable legal injuries.²⁷⁰ To bridge this gap, IHRL must move toward structural scrutiny that prioritizes the preservation of mental integrity and the autonomous self-development of the individual over mere procedural compliance.²⁷¹ Such scrutiny would not abandon existing proportionality doctrine, but would require its principles of legality, necessity, safeguards, and fair balance to account for cumulative psychological harms and structurally embedded interferences with the inner sphere.²⁷²

4.4.4 Operationalizing Human Rights Due Diligence as a Tool for Structural Scrutiny

The transition toward binding AI governance is anchored in the requirement that States and private enterprises move beyond voluntary ethics and operationalize the UNGP on Business and Human Rights.²⁷³ A central component of this responsibility is HRDD, which requires an ongoing process to identify, prevent, mitigate, and account for adverse human rights impacts. In the context of AI systems, HRDD is particularly significant because the relevant harms often do not arise only after a discrete violation has occurred.²⁷⁴ Rather, they may emerge gradually through the design,

²⁶⁵ *ibid.*, p. 218.

²⁶⁶ Teo, (n 2)

²⁶⁷ Cohen, (n 55) p. 1911.

²⁶⁸ Teo, (n 2)

²⁶⁹ Teo, (n 165) p. 226.

²⁷⁰ *ibid.*, p. 218.

²⁷¹ *ibid.*, p. 216.

²⁷² *ibid.*, p. 216; Warso, (n 13) p. 79.

²⁷³ A/HRC/59/53

²⁷⁴ Teo, (n 165) p. 218.

procurement, deployment, and continued operation of systems that enable exploitative surveillance, behavioral manipulation, or the classification of individuals through opaque inferential processes.²⁷⁵

Recent work of the UN Working Group on Business and Human Rights confirms the importance of this preventive approach. The Working Group has warned that States and businesses are procuring and deploying AI systems without sufficiently robust HRDD, thereby creating risks of exploitative worker surveillance, biased decision-making, and manipulation of consumer behavior.²⁷⁶ This is directly relevant to the argument advanced in this thesis. If AI harms are cumulative, structurally embedded, and psychologically mediated, then rights protection cannot depend solely upon downstream remedies after harm becomes visible.²⁷⁷ HRDD becomes significant because it shifts scrutiny upstream, requiring States and businesses to assess the human rights implications of AI systems before and throughout their deployment.²⁷⁸

First, HRDD must operate across the full AI lifecycle.²⁷⁹ It cannot be reduced to a one-time compliance exercise conducted at the point of procurement or design. This lifecycle approach is legally important because many of the harms examined in this thesis, including chilling effects, anticipatory conformity, and behavioral modulation, accumulate gradually rather than manifesting as a single temporally proximate injury.²⁸⁰ A system that appears proportionate at the moment of deployment may nevertheless become rights-intrusive over time as it continuously observes, classifies, and reshapes the informational environment of affected individuals.²⁸¹ Continuous HRDD therefore provides a mechanism through which the slow violence of AI systems can be monitored, reassessed, and mitigated before it hardens into structural harm.

Second, HRDD may address the epistemic asymmetry created by black-box algorithmic governance. As the analysis of SyRI demonstrated, opacity does not merely create a technical problem; it weakens the individual's ability to understand, challenge, and contest the basis upon which they are assessed. By requiring public authorities and businesses to disclose the capabilities, limitations, and human rights impacts of AI systems, HRDD can function as

²⁷⁵ UN Working Group on Business and Human Rights, *Artificial Intelligence Procurement and Deployment: Ensuring Alignment with the Guiding Principles on Business and Human Rights* (2025) 2; Office of the United Nations High Commissioner for Human Rights, *The Right to Privacy in the Digital Age* UN Doc A/HRC/48/31 (2021) para 33.

²⁷⁶ *ibid.*

²⁷⁷ Teo, (n 165) p. 218.

²⁷⁸ *ibid.*

²⁷⁹ A/HRC/48/31

²⁸⁰ Teo, (n 165) p. 218.

²⁸¹ Council of Europe Committee of Ministers, *Recommendation CM/Rec(2020)1 to member States on the human rights impacts of algorithmic systems* (adopted 8 April 2020 at the 1373rd meeting of the Ministers' Deputies) CM/Rec(2020)1, para 10; Council of Europe Committee of Ministers, *Declaration by the Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes* (adopted 13 February 2019 at the 1337th meeting of the Ministers' Deputies) Decl(13/02/2019)1, para 7.

a safeguard against the creation of artificially immutable risk profiles.²⁸² Such disclosure is not valuable merely as informational transparency.²⁸³ Its legal significance lies in whether it enables meaningful scrutiny, contestation, and accountability. In this sense, HRDD supports the proportionality requirement by requiring duty-bearers to demonstrate that an AI system is not only useful, but also necessary, accurate, and accompanied by adequate safeguards.²⁸⁴

Third, HRDD is particularly important where AI systems operate through behavioral manipulation or exploit situations of vulnerability.²⁸⁵ Systems that rely upon personalized choice architectures, hypernudging, or emotional inference may interfere with privacy not by exposing personal information alone, but by shaping the conditions under which individuals make decisions. A meaningful HRDD process must therefore assess whether the design of an AI system exploits vulnerability, bypasses rational agency, or creates dependency in ways that undermine mental autonomy and dignity.²⁸⁶ This is especially significant in contexts such as workplace monitoring, social protection, healthcare, mental health support, and consumer platforms, where individuals may be structurally dependent upon the system or unable to meaningfully refuse participation.²⁸⁷

Accordingly, HRDD can serve as an important tool for bridging the proportionality gap identified in this thesis. It requires duty-bearers to move beyond abstract claims of efficiency or innovation and provide evidence of necessity, accuracy, proportionality, and safeguards before and during deployment.²⁸⁸ However, HRDD should not be understood as sufficient in itself.²⁸⁹ If reduced to a procedural checklist, it risks reproducing the same limitations identified in existing proportionality analysis. Its value lies in its capacity to operationalize structural scrutiny by forcing States and businesses to evaluate how AI systems affect autonomy, behavioral freedom, informational self-determination, and the integrity of the inner sphere over time.²⁹⁰

Therefore, HRDD becomes legally meaningful only when coupled with substantive safeguards. These may include independent oversight, meaningful contestability, restrictions on manipulative design, protection against exploitative surveillance, and a right to human review in high-risk contexts. Such safeguards ensure that the logic of data extraction and prediction does not override the structural conditions necessary for autonomous self-development. In this way, HRDD does not replace proportionality analysis, but strengthens it by making visible the cumulative

²⁸² UN Working Group on Business and Human Rights, (n 257), p. 4

²⁸³ Wachter, (n 120) pp. 45-46.

²⁸⁴ A/HRC/48/31, para. 54.

²⁸⁵ A/HRC/59/53

²⁸⁶ Teo, (n 2), pp. 4-5

²⁸⁷ A/HRC/48/31, paras. 31-33.

²⁸⁸ Warso, (n 35) p. 75.

²⁸⁹ Teo, (n 165) p. 216.

²⁹⁰ *ibid.*; Teo, (n 2), p. 13.

and structurally embedded harms that conventional review often fails to capture.²⁹¹

²⁹¹ Teo, (n 165) pp. 216-218

5 Conclusion

This thesis has argued that the proliferation of AI-driven systems presents a fundamental challenge to the protection of privacy under Article 17 ICCPR and Article 8 ECHR. The challenge does not arise only because AI systems collect, process, or disclose personal data at scale. Rather, it arises because such systems increasingly interfere with the individual through more structurally embedded mechanisms of behavioural inference, opacity, and environmental modulation. The core interference identified throughout this research is therefore not merely an external invasion of privacy, but a gradual reconfiguration of the conditions under which individuals think, decide, behave, and develop identity. Through psychological pressure, anticipatory conformity, and persistent algorithmic evaluation, AI systems may impair autonomy, identity formation, and the foundational integrity of the “inner sphere.”

The central finding of this thesis is that existing proportionality doctrine already contains the necessary legal mechanisms to respond to these harms. The principles of legality, foreseeability, necessity, safeguards, fair balance, and positive obligations provide a doctrinal foundation through which AI-mediated interferences with privacy can be scrutinised. However, the current application of proportionality remains limited by what this thesis has described as a “utility blind spot” and by continued reliance upon a “salient harm” model. Courts and regulatory frameworks are often more comfortable addressing harms that are visible, immediate, measurable, or procedurally identifiable, such as unlawful data processing, lack of notice, insufficient transparency, or discriminatory outputs. Yet AI-mediated harms frequently operate differently. They may accumulate gradually, remain psychologically embedded, and manifest through the erosion of agency rather than through a single discrete injury. This creates a “proportionality gap” in which the law may identify the technical “black box” while still failing to fully cognise the “slow violence” of algorithmic herding and the attrition of the *forum internum*.

The thesis developed this argument in four stages. First, the doctrinal analysis established that privacy under Article 17 ICCPR and Article 8 ECHR cannot be reduced to informational secrecy. Rather, privacy functions as a structural condition for autonomous self-development, protecting the space within which individuals form beliefs, test ideas, construct identity, and engage in boundary management away from excessive observation. This doctrinal foundation is essential because AI-mediated interferences often do not target privacy only as data control, but as the broader condition for independent thought and self-formation.

Second, the analysis of AI harms demonstrated how behavioural inference, digital phenotyping, predictive analytics, and emotional inference transform privacy interference into a process of continuous evaluation. Unlike traditional surveillance, which often involved identifiable acts of observation or disclosure, AI systems may infer latent characteristics, vulnerabilities,

emotional states, and behavioural tendencies from ordinary digital traces. These systems may then shape informational environments in ways that exploit vulnerability and induce conformity effects that bypass rational agency. In this sense, the harm is not simply that individuals are watched, but that they are gradually governed through prediction, classification, and modulation. This is why the thesis framed AI harm as cumulative, relational, and structurally embedded rather than merely informational.

Third, the case studies of *SyRI* and *Bărbulescu v Romania* illustrated both the strength and the limitation of existing doctrine. *SyRI* demonstrated that proportionality analysis can recognise opacity, informational asymmetry, lack of foreseeability, insufficient safeguards, and the dangers of state algorithmic governance. It therefore supports the argument that States bear a heightened responsibility when deploying sophisticated technologies capable of extensive interference with private life. However, the judgment remained primarily focused on procedural opacity, data minimisation, discriminatory risk, and safeguards. It did not fully articulate the psychological effects of living under persistent algorithmic suspicion or the behavioural consequences of being governed through opaque risk profiles. Similarly, *Bărbulescu* showed that Article 8 ECHR imposes positive obligations upon States to protect individuals from disproportionate monitoring by private actors. Yet the Court's reasoning remained heavily oriented toward notice, procedural safeguards, and the balancing of employer and employee interests. It did not fully confront the deeper structural harm produced by persistent workplace surveillance: the pressure to perform an idealised professional persona under the algorithmic gaze.

Fourth, the analysis of emerging AI governance showed that States and international institutions are increasingly moving toward binding human-rights-based regulation. Instruments such as the EU AI Act, HRDD frameworks, transparency registers, and FRIAs represent important attempts to respond to the risks of opacity, discrimination, manipulation, and insufficient oversight. However, these developments also reveal the continuing limits of proceduralism. If HRDD, transparency, and impact assessments are treated as technical compliance exercises, they risk becoming “check-box” mechanisms rather than substantive tools for protecting autonomy. Their value lies not in their formal existence, but in whether they require States and private actors to confront the structural effects of AI systems upon behavioural freedom, mental autonomy, and the inner personal core.

In conclusion, this thesis contends that the proportionality test must move toward a model of structural scrutiny capable of evaluating the “conditions of possibility” for autonomous self-development. This does not require abandoning existing doctrine. Rather, it requires applying legality, necessity, safeguards, fair balance, and positive obligations in a manner capable of recognising cumulative psychological and structurally embedded harms as legally relevant interferences with privacy. States may therefore fail to comply with their privacy obligations where they deploy, permit, or

inadequately regulate AI systems without sufficiently considering the harms produced by persistent algorithmic observation, inference, and modulation. The violation does not lie merely in the existence of AI systems themselves, but in the failure to ensure that their design, deployment, oversight, and regulation adequately protect the *forum internum* against gradual erosion.

Substantive safeguards such as meaningful oversight, contestability, human review in high-risk contexts, and restrictions on manipulative design should therefore be understood as mechanisms that strengthen proportionality analysis rather than replace it. They are necessary because privacy in the AI era cannot be protected only by requiring notice, transparency, or formal legality. It must also protect individuals against systems that reshape the very conditions under which they think, choose, and become themselves.

6 Bibliography

Academic Literature

Cohen JE, ‘Surveillance vs Privacy: Effects and Implications’ in David Gray and Stephen E Henderson (eds), *The Cambridge Handbook of Surveillance Law* (Cambridge University Press 2017)

Cohen JE, ‘What Privacy Is For’ (2013) 126 *Harvard Law Review* 1904

Cosgrove L and others, ‘Digital Phenotyping and Digital Psychotropic Drugs: Mental Health Surveillance Tools That Threaten Human Rights’ (2020) 22(2) *Health and Human Rights Journal* 33

Istace Z, ‘Protecting the Mental Realm: What Does Human Rights Law Bring to the Table?’ (2023) 41(4) *Netherlands Quarterly of Human Rights* 214

Joseph S and Castan M, *The International Covenant on Civil and Political Rights: Cases, Materials, and Commentary* (3rd edn, Oxford University Press 2013)

Molbæk-Steensig H and Scheinin M, ‘Human Rights and Artificial Intelligence in Healthcare-related Settings’ in Tine Hanrieder, Aadya Sharma and Paul B Martins (eds), *The Future of Human Rights Governance* (Cambridge University Press 2023)

Penney JW, ‘Understanding Chilling Effects’ (2022) 106(3) *Minnesota Law Review* 1451

Richards NM, ‘Intellectual Privacy’ (2008) 87 *Texas Law Review* 387

Richards NM, *Intellectual Privacy: Rethinking Civil Liberties in the Digital Age* (Oxford University Press 2015)

Steenstra I and others, ‘Assessing Risks of Large Language Models in Mental Health Support: A Framework for Automated Clinical AI Red Teaming’ (2026) <<https://doi.org/10.48550/arXiv.2602.19948>> accessed 3 May 2026

Teo SA, ‘How Artificial Intelligence Systems Challenge the Conceptual Foundations of the Human Rights Legal Framework’ (2022) 40(1) *Nordic Journal of Human Rights* 216

Teo SA, 'Human Dignity and AI: Mapping the Contours and Utility of Human Dignity in Addressing Challenges Presented by AI' (2023) 15(1) *Law, Innovation and Technology* 257

Teo SA, 'How to Think About Freedom of Thought and Opinion in the Age of AI' (2024) 53 *Computer Law & Security Review* 105969

Teo SA, 'Artificial Intelligence and Its "Slow Violence" to Human Rights' (2025) 5(3) *AI and Ethics* 2265

Teo SA, 'Artificial Intelligence, Human Vulnerability and Multi-Level Resilience' (2025) 57 *Computer Law & Security Review* 106134

Tobin J and Field SM, 'Article 16: The Right to Privacy, Family, Home, Correspondence, Honour and Reputation' in Jonathan Todres and Shani M King (eds), *The Oxford Handbook of Children's Rights Law* (Oxford University Press 2020)

Wachter S, 'The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law' (2022–2023) 97 *Tulane Law Review*

Warso Z, 'Human Rights Requirements for Person-Based Predictive Policing: Lessons from Selected ECtHR Case Law and Its Limits' (2022) *Technology and Regulation* 75

Jurisprudence

Bărbulescu v Romania App no 61496/08 (ECtHR, Grand Chamber, 5 September 2017)

Big Brother Watch and Others v United Kingdom App nos 58170/13, 62322/14 and 24960/15 (ECtHR, Grand Chamber, 25 May 2021)

Handyside v United Kingdom (1976) 1 EHRR 737

NJCM c.s. and others v De Staat der Nederlanden District Court of The Hague, 5 February 2020, ECLI:NL:RBDHA:2020:1878

Legal Instruments

Convention for the Protection of Human Rights and Fundamental Freedoms 1950

International Covenant on Civil and Political Rights 1966

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L 2024/1689

Official Documents

Council of Europe Committee of Ministers, *Declaration by the Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes* Decl(13/02/2019)1 (adopted 13 February 2019)

Council of Europe Committee of Ministers, *Recommendation CM/Rec(2020)1 to Member States on the Human Rights Impacts of Algorithmic Systems* CM/Rec(2020)1 (adopted 8 April 2020)

Human Rights Committee, *General Comment No 16: Article 17: Right to Privacy* (8 April 1988)

Human Rights Committee, *General Comment No 27: Article 12: Freedom of Movement* UN Doc CCPR/C/21/Rev.1/Add.9 (2 November 1999)

Office of the United Nations High Commissioner for Human Rights, *Guiding Principles on Business and Human Rights: Implementing the United Nations “Protect, Respect and Remedy” Framework* (United Nations 2011)

UN Human Rights Council, *The Right to Privacy in the Digital Age: Report of the United Nations High Commissioner for Human Rights* UN Doc A/HRC/39/29 (3 August 2018)

UN Human Rights Council, *The Right to Privacy in the Digital Age: Report of the United Nations High Commissioner for Human Rights* UN Doc A/HRC/48/31 (13 September 2021)

UN Working Group on Business and Human Rights, *The Practical Application of the Guiding Principles on Business and Human Rights to the Activities of Technology Companies* UN Doc A/HRC/59/53 (2025)

UN Working Group on Business and Human Rights, *Artificial Intelligence Procurement and Deployment: Ensuring Alignment with the Guiding Principles on Business and Human Rights* (2025)

United Nations, *Global Principles for Information Integrity: Recommendations for Multi-stakeholder Action* (United Nations 2024)

Other Documentation and Materials

Shany Y, *The Need for and Feasibility of an International AI Bill of Human Rights* (White Paper, Institute for Ethics in AI, University of Oxford, November 2025)

Article 5: Prohibited AI Practices (EU Artificial Intelligence Act)
<https://artificialintelligenceact.eu/article/5/> accessed 3 May 2026

Acknowledgement of AI Use

ChatGPT was used as a supporting tool for:

1. translation and comprehension support, particularly for understanding complex legal sentences, academic texts, and legal terminology;
2. language enhancement, including grammar and clarity;
3. research orientation, including identifying potentially relevant news, reports related to AI, privacy, mental autonomy, and human rights;

All sources and legal materials identified with the assistance of ChatGPT were independently verified and evaluated. ChatGPT was not used as a substitute for independent legal analysis, source assessment, or academic judgment.