

RESEARCH PAPER

The Synthetic Identity Threat

Governance Frameworks for AI-Generated Personas

A Digital Dignity Institute comparative analysis of eight jurisdictions, and a proposed minimum standard for the disclosure, liability, and enforcement architecture that synthetic personas now require.

*May 2026. DDI Research Paper Series. The Digital Dignity Institute of California, San Francisco.
Issued institutionally. Not attributed to individual authors.*

Abstract

As artificial intelligence enables the creation of convincing synthetic identities, meaning false personas that are indistinguishable from real individuals, existing legal frameworks are inadequate to the harms that follow. This brief examines regulatory approaches across eight jurisdictions and proposes minimum standards for synthetic identity disclosure, liability, and enforcement.

The analysis proceeds from a structural observation. Identity law across virtually every jurisdiction is built on the assumption that identity harms involve two natural persons: one whose identity is taken and one who takes it. Synthetic identity breaks that assumption. The persona is fabricated rather than appropriated, which means there is no complainant with standing under identity theft statutes, no plaintiff with a publicity or defamation claim, and no clear locus of harm on which enforcement can rest. The result is a governance vacuum in which the most scalable form of identity deception is also the least regulated.

We survey the three regulatory models currently in use, namely disclosure, prohibition, and liability, and assess each against the conditions under which it can plausibly work. We conclude that disclosure regimes are necessary but insufficient, that prohibition regimes are workable only within narrowly defined contexts, and that civil liability with a rebuttable presumption against creators and deployers offers the most promising path to deterrence. We propose four minimum standards and eight policy recommendations for legislatures, regulators, and platforms.

KEY FINDINGS

- 01.** Existing identity fraud and impersonation laws do not cover synthetic identities. There is no victim whose identity has been stolen.
- 02.** Disclosure-based regimes are necessary but insufficient. Labelling requirements are easily circumvented and do not address harms already caused.
- 03.** Civil liability regimes with rebuttable presumptions offer the most promising path to deterrence.
- 04.** Platform-level detection obligations are essential. Individual enforcement cannot address synthetic identity at scale.
- 05.** International coordination is required. Synthetic identity operations are inherently cross-border.

Contents

1. Introduction: The Synthetic Identity Threat
2. Defining Synthetic Identity
3. The Structural Inadequacy of Existing Law
4. Regulatory Approaches Across Eight Jurisdictions
5. Assessing the Three Models
6. Minimum Standards for Adequate Governance
7. Objections and Responses
8. Implementation Pathways
9. Policy Recommendations
10. Conclusion
 - Annex A: Comparative Coding Framework
 - Annex B: Model Statutory Language
 - Sources and Further Reading

1. Introduction: The Synthetic Identity Threat

1.1 What has changed

Advances in generative artificial intelligence have made it possible to create synthetic identities, meaning false personas with convincing photographs, plausible biographical histories, and coherent online presences, at scale and at negligible cost. These capabilities are being deployed for fraud, influence operations, harassment, and the systematic erosion of epistemic trust.

None of the individual components of a synthetic identity is new. Fabricated personas have been a staple of intelligence tradecraft, undercover policing, and confidence fraud for as long as those practices have existed. What has changed is the cost curve and the quality ceiling. A persona that in 2015 required a stolen photograph, a hand-written backstory, and weeks of manual posting to appear plausible can now be generated in minutes: a face that belongs to no living person, a résumé consistent with that face, a voice that matches the résumé, a posting history that matches the voice, and a social graph assembled from other personas produced the same way. The marginal cost of the thousandth persona is close to the marginal cost of the first.

Three properties of this shift matter for governance.

The first is **volumetric**. Deception that once scaled linearly with human effort now scales with compute. This changes the enforcement calculus fundamentally, because remedies designed around individual bad actors cannot absorb adversaries who operate at population scale.

The second is **qualitative**. Synthetic personas are no longer detectable by the heuristics that ordinary people use, such as implausible photographs, inconsistent details, or awkward language. The detection burden has moved from the human eye to forensic tooling that most individuals and many institutions do not possess.

The third is **compositional**. Synthetic identities are increasingly multimodal and persistent. A persona is not a single fake image but a durable entity with an email address, a payment method, a professional profile, a publication record, a voice, and a history of interaction. Persistence is what converts a fake into a relationship, and relationships are what make fraud, influence, and harassment effective.

Taken together, these properties describe a category of harm that legal systems have not previously had to address: deception at scale by entities that do not exist, in contexts where the existence of the counterparty was an unstated assumption of the entire transaction.

1.2 Why existing law fails

Existing legal frameworks, designed to address identity theft and impersonation of real individuals, are structurally inadequate for synthetic identity. The inadequacy is not a matter of drafting oversight that can be corrected by amendment. It follows from the architecture of identity law itself.

Identity theft statutes protect a person whose identifying information has been appropriated. Impersonation offences protect a person whose likeness or persona has been assumed. Rights of publicity protect the commercial value of a real individual's identity. Defamation protects reputation, which is a property of real people and legal persons. Every one of these instruments presumes a claimant.

A synthetic identity has no claimant. There is no victim whose identity has been stolen, no real person who has been impersonated, and no clear locus of harm. There are, of course, people who are harmed: the investor defrauded by a fictitious fund manager, the electorate manipulated by a network of fictitious constituents, the researcher whose field is polluted by fictitious peers. But their harms are diffuse, downstream, and difficult to trace to a specific act by a specific actor. The law is well equipped to remedy a stolen identity and poorly equipped to remedy a manufactured one.

This paper builds on two analytical concepts that the Institute has developed across its previous work and that apply with particular force here.

The **recourse deficit** describes the gap between the harms that a technology produces and the remedies that are

actually available to those who suffer them. A recourse deficit is not the absence of law in the abstract. It is the absence of a route from an identifiable injury to an identifiable remedy that an ordinary person can actually travel.

The **discovery burden** describes the cost, borne by the harmed party, of establishing that a harm occurred at all. In most legal contexts this burden is trivial: a person knows they were defrauded, defamed, or dismissed. Synthetic identity inverts this. The defining feature of a successful synthetic persona is that the person interacting with it does not know it is synthetic, and may never know. Where the discovery burden is high enough, the recourse deficit becomes total, because a remedy that is never invoked is functionally identical to a remedy that does not exist.

Section 3 develops both concepts in relation to specific bodies of law. They are, throughout this paper, the criteria against which each regulatory model is assessed. A regime that assigns liability but leaves the discovery burden with the individual will not deliver recourse. A regime that mandates disclosure but provides no remedy for non-disclosure will not deliver deterrence.

1.3 Scope, method, and limitations

This paper examines regulatory approaches to synthetic identity across eight jurisdictions: the European Union, the United States, the United Kingdom, China, South Korea, Canada, Brazil, and India. The selection is intended to span legal traditions (common law, civil law, and mixed), regulatory philosophies (precautionary, permissive, and state-directed), and stages of legislative development (comprehensive statute in force, statute pending, and sectoral instruments only). It is not a representative global sample, and the omission of the African Union, the Gulf states, Japan, and Australia reflects space rather than significance.

The method is doctrinal and comparative. For each jurisdiction we identify the instruments that bear on synthetic identity, whether or not they were drafted with it in mind, and code them against a common framework set out in Annex A. That framework asks six questions of each regime: what conduct is regulated, who bears the obligation, what triggers it, what remedy attaches, who may invoke the remedy, and where the discovery burden sits. Coding was performed against primary sources, meaning statutory text, implementing regulations, and published regulatory guidance, with secondary commentary used only for interpretive context.

Three limitations should be stated plainly.

First, the field is moving faster than the literature. Statements of legal position are current as of May 2026 and several of the instruments discussed are in transitional application periods, under judicial challenge, or awaiting implementing regulation. Readers should verify current status before relying on any jurisdictional description for operational purposes.

Second, this paper does not conduct original empirical research on the prevalence of synthetic identity. It draws on published platform transparency reporting, regulatory enforcement records, and the security research literature. Prevalence estimates in this domain are weak, and we treat them as indicative rather than probative.

Third, the analysis is deliberately confined to synthetic identity as defined in Section 2. It does not attempt a general theory of synthetic media governance, and it treats deepfakes of real individuals as a distinct problem with a distinct and better-developed legal answer.

2. Defining Synthetic Identity

2.1 A working definition

Regulatory failure in this domain frequently begins with definitional imprecision. Instruments drafted to address “AI-generated content” sweep in satire, art, accessibility tooling, and legitimate pseudonymity, while instruments drafted to address “deepfakes” reach only depictions of real people and miss fabricated ones entirely. A workable definition needs to be narrow enough to survive constitutional scrutiny and broad enough to capture the conduct that actually causes harm.

We propose the following.

A **synthetic identity** is a persona presented as a natural person that (i) has no corresponding natural person, (ii) is constituted wholly or substantially by machine-generated artefacts across one or more modalities, (iii) persists across time or interactions, and (iv) is deployed in a context where the audience would reasonably assume they are engaging with a natural person.

Each element does work.

No corresponding natural person distinguishes synthetic identity from deepfakes and impersonation, which involve real referents and are governed by other law.

Machine-generated constitution distinguishes it from ordinary fabrication, such as a purely textual pen name, and locates the regulatory interest in the capability that changed the cost curve.

Persistence distinguishes an identity from an artefact. A single generated image is a synthetic image. A generated image attached to a name, a history, and a pattern of interaction is a synthetic identity, and persistence is the property that enables trust to accrue.

Contextual assumption of humanity is the element that carries the harm. A persona that announces itself as an AI system in a setting where that is expected, such as a labelled customer service agent or a character in a game, is not deceiving anyone. The same persona presented as a constituent, a customer, a peer reviewer, or a romantic prospect is.

2.2 What synthetic identity is not

Precision here protects legitimate practice, and the categories below should be expressly carved out of any instrument built on this definition.

Pseudonymity. A real person writing under a chosen name is exercising a protected practice with a long democratic history. Pseudonymity conceals which natural person is speaking; synthetic identity asserts that a natural person exists where none does. The distinction is not merely conceptual: pseudonymity leaves a real agent who can be held accountable, and synthetic identity does not.

Avatars and declared virtual personas. Digital representations of real people, and virtual figures that are openly presented as such, including virtual influencers who disclose their nature, fall outside the definition because element (iv) is not satisfied.

Fiction and satire. Characters in narrative works are not deployed in contexts where the audience assumes a natural person. Where satire adopts the form of a real or apparently real persona, the analysis turns on context and on established expressive protections, and we address this directly in Section 7.

Synthetic identity fraud in the credit sense. The financial crime literature uses “synthetic identity fraud” to describe the combination of real identifying data, typically a genuine identification number, with fabricated details to create a credit file. This is a serious and well-documented harm, but it is a documentary offence with an existing statutory home. Our subject is the persona rather than the file, and the two overlap without being coextensive.

Deepfakes of real individuals. Non-consensual synthetic depictions of real people are a grave harm and are the subject of substantial recent legislative activity across our sample. They are excluded here precisely because that legislative activity exists. Synthetic identity is the residual category that the deepfake response leaves untouched.

2.3 A typology of deployment

Synthetic identities are protean instruments rather than ends. The harms they produce depend on deployment context, and effective regulation is likely to be context-specific for reasons developed in Section 5. We identify seven deployment categories.

Financial fraud. Fabricated fund managers, advisers, brokers, and counterparties. The persona supplies the credibility that the underlying scheme lacks. Long-form relationship fraud, including the investment scams commonly described as pig butchering, depends on sustained interaction with a persona that appears to be a real person with a real life, and generative tooling has reduced both the labour cost and the linguistic barriers that previously constrained these operations.

Influence operations. Networks of fabricated constituents, activists, journalists, and experts deployed to simulate public opinion. The harm here is not that any individual is deceived about any particular fact, but that the perceived distribution of opinion is manipulated. This is the category in which the absence of an individual victim is most acute and in which the collective harm is most severe.

Harassment and coordinated abuse. Synthetic personas allow a single actor to simulate a crowd. Targets experience what appears to be widespread hostility and cannot distinguish organic disapproval from manufactured pressure. The discovery burden falls on the person least able to bear it.

Market and reputational manipulation. Fabricated reviewers, customers, employees, and analysts. Consumer protection law is the closest existing fit and its limitations are examined in Section 3.3.

Epistemic pollution. Fabricated authors, peer reviewers, expert commentators, and sources. Where synthetic personas enter the systems by which societies produce and validate knowledge, including journalism, scholarship, and standards development, the damage is to the validation mechanism rather than to any individual participant.

Credential and access fraud. Synthetic personas used to obtain employment, particularly remote employment with network access, to secure professional accreditation, or to pass identity verification. This category has attracted the most attention from security agencies and produces the clearest evidentiary trail.

Relational exploitation. Romance fraud, grief exploitation, and the cultivation of emotional dependency on a persona that does not exist. This category produces harms that are severe, deeply personal, and poorly captured by any of the frameworks surveyed here, because the loss is relational and the legal system recognises primarily financial injury.

The categories are not exclusive. A single operation frequently spans several, and the highest-consequence operations typically combine relational cultivation with financial or influence objectives. Any governance framework built around a single category will be evaded by operations that straddle two.

3. The Structural Inadequacy of Existing Law

This section examines the five bodies of law most likely to be invoked against synthetic identity harms and identifies, for each, the structural feature that defeats the claim. The purpose is to establish that the governance gap is architectural rather than incidental, and therefore that it cannot be closed by extending existing instruments at the margins.

3.1 Identity theft and fraud statutes: the missing victim

Identity crime statutes across our sample share a common structure. They prohibit obtaining, possessing, transferring, or using identifying information *of another person* without lawful authority. The qualifying phrase is doing the work, and it is fatal.

In United States federal law, the aggravated identity theft provision reaches the use of a means of identification of another person. Where the identification belongs to no one, the element is not made out. Courts have consistently required a real person behind the identifying information, and prosecutors have accordingly charged synthetic identity conduct under wire fraud provisions instead, which reach the scheme rather than the identity. The same pattern holds in the United Kingdom, where the Fraud Act 2006 offence of fraud by false representation reaches the representation and not the persona, and in the civil law jurisdictions in our sample, where identity offences are located in provisions protecting personality rights that only real persons hold.

The practical consequence is that synthetic identity is prosecutable only where it is instrumental to a separately recognised offence, typically fraud. This has three effects. It leaves non-financial harms, particularly influence and harassment, without a criminal route. It requires proof of the downstream scheme, which is harder to establish than the identity conduct itself. And it produces no deterrent against the creation and maintenance of persona infrastructure, which is the part of the operation that scales.

3.2 Publicity, personality, and defamation: the missing plaintiff

Rights of publicity, personality rights in the civil law tradition, and defamation all protect interests attached to a specific individual. A fabricated persona has no dignity interest, no reputational interest, and no commercial identity value, because it has no self to which those interests could attach.

Two partial exceptions deserve mention.

Where a synthetic persona is constructed from training data or reference images derived from identifiable individuals, and where the output bears a recognisable resemblance to a real person, a publicity or personality claim may survive. The evidentiary threshold is high and the analysis is unstable. As generation techniques improve at producing outputs that resemble no particular individual, this route narrows further.

Where a synthetic persona defames a real person, the person defamed has a claim, but against whom? The persona cannot be sued. The claim must reach an operator whose identity is typically unknown, which returns us to the discovery burden. Defamation law assumes a locatable defendant, and synthetic identity operations are designed to ensure there is not one.

3.3 Consumer protection: the closest fit and its limits

Consumer protection law is the most promising of the existing frameworks, because it regulates deceptive conduct without requiring an identified victim. Prohibitions on unfair and deceptive acts and practices, on misleading commercial

communications, and on fake endorsements can reach synthetic personas deployed in commercial contexts.

In the United States, the Federal Trade Commission's rule on government and business impersonation, finalised in 2024, established a strong precedent by attaching liability to impersonation conduct itself and creating a route to monetary relief. The Commission's contemporaneous proposal to extend the rule to the impersonation of individuals, together with its guidance on fake reviews and testimonials, moved further in the right direction. The European framework achieves comparable coverage through the Unfair Commercial Practices Directive and its prohibition on falsely representing oneself as a consumer.

Three limits constrain this route.

Commercial confinement. Consumer protection instruments reach commercial practice. Influence operations, harassment campaigns, and epistemic pollution are frequently non-commercial, and the most consequential synthetic identity harms fall outside the jurisdictional grant.

Regulator-dependence. These are public enforcement regimes with limited or absent private rights of action. Enforcement capacity is finite and case selection favours large, provable, domestically located schemes. Synthetic identity operations are numerous, distributed, and typically foreign.

Individual framing. The instruments target the deceptive act rather than the infrastructure that produces deceptive acts at scale. Enforcement against a persona is enforcement against the least valuable component of the operation.

3.4 Intermediary liability and the platform question

Platforms host the personas, and platforms possess the signals, including registration patterns, device fingerprints, behavioural telemetry, and network structure, that make detection tractable at scale. They are, for practical purposes, the only actors positioned to address the problem systemically. Yet across our sample, platform obligations in respect of synthetic identity are either absent or framed as content moderation duties that fit the problem poorly.

Intermediary liability protections, whether the United States approach of broad immunity for third-party content or the European approach of conditional immunity contingent on notice and action, were designed around content. Synthetic identity is not principally a content problem. The relevant artefact is the account, and the relevant question is not whether a particular post is unlawful but whether an entity exists. Frameworks organised around unlawful content are poorly suited to an inquiry organised around entity authenticity, and the mismatch has left the most capable actors with the fewest defined obligations.

3.5 The recourse deficit and the discovery burden applied

Bringing the analysis together. For each body of law, the following table locates the point of failure and identifies where the discovery burden sits.

Body of law	Structural failure	Discovery burden sits with
Identity theft statutes	No “other person” whose identity was taken	Investigating authority, triggered only by a reported downstream loss
Impersonation offences	No real persona assumed	Individual complainant
Publicity and personality rights	No claimant with a protected interest	Individual, if resemblance is noticed
Defamation	No locatable defendant	Individual, at high evidentiary cost
Consumer protection	Commercial contexts only; public enforcement only	Regulator, with limited capacity
Intermediary liability	Content-shaped duties for an entity-shaped problem	Platform, but with no obligation to look

The pattern is consistent. In every case where a remedy theoretically exists, the burden of discovering that a harm occurred falls on a party that either lacks the tools to discover it or lacks the incentive to look. The recourse deficit in synthetic identity is therefore not primarily a deficit of substantive rules. It is a deficit of detection, attribution, and standing, and any framework that adds rules without addressing those three will produce paper rights.

4. Regulatory Approaches Across Eight Jurisdictions

4.1 Method and models

This brief examines regulatory approaches to synthetic identity across eight jurisdictions, identifying three broad models.

Disclosure-based regimes require AI-generated content, personas, or interactions to be labelled or otherwise revealed. The regulatory theory is informational: audiences equipped with accurate provenance information can protect themselves.

Prohibition-based regimes restrict the creation or deployment of synthetic identities in specified contexts, typically elections, financial services, or sexual imagery. The regulatory theory is categorical: certain deployments are harmful enough that disclosure is not an adequate cure.

Liability-based regimes assign responsibility for harms caused by synthetic identities to their creators or deployers. The regulatory theory is deterrent: internalising the cost of harm changes the economics of production.

The models are not mutually exclusive and most jurisdictions in our sample combine elements of two or three, usually without an integrating theory. What follows assesses each jurisdiction against the coding framework in Annex A, with attention throughout to where the discovery burden is placed.

4.2 European Union

The European Union has the most developed framework in our sample, and it is a disclosure regime with prohibition elements at the edges.

The AI Act, in force since August 2024, establishes transparency obligations in Article 50 that bear directly on synthetic identity. Providers of systems intended to interact directly with natural persons must ensure that those persons are informed they are interacting with an AI system unless it is obvious from the context. Deployers of systems that

generate or manipulate content constituting a deep fake must disclose that the content has been artificially generated or manipulated. Providers of generative systems must ensure outputs are marked in a machine-readable format detectable as artificially generated. These obligations apply from August 2026, with the associated codes of practice and harmonised standards in development through the transitional period.

Article 50 is a significant advance and it does not close the synthetic identity gap for three reasons. The interaction disclosure obligation attaches to systems that interact with natural persons, which reaches conversational agents but not persona infrastructure created by one system and deployed manually elsewhere. The deep fake disclosure obligation is defined by reference to resemblance to existing persons, objects, or events, which sits awkwardly with personas that resemble nobody. And the marking obligation applies to providers, meaning that content generated by open-weight or offshore systems, or stripped of provenance metadata downstream, falls outside effective coverage.

The Digital Services Act supplies the platform layer. Very large online platforms must assess and mitigate systemic risks, including risks to civic discourse and electoral processes and the dissemination of inauthentic content, and must submit to independent audit. Inauthentic account behaviour is within scope of the risk assessment obligation, and the Commission's enforcement activity has made coordinated inauthentic behaviour an explicit area of supervisory interest. The DSA therefore comes closer than any other instrument in our sample to imposing a systemic detection duty, though it does so through the indirect mechanism of risk management rather than a direct obligation to identify synthetic accounts.

The General Data Protection Regulation is relevant chiefly by its absence. A wholly synthetic persona is not a natural person and its attributes are not personal data, so the Regulation does not apply to the persona. It applies to the individuals the persona targets and to the data used to generate it, which produces the counterintuitive result that Europe's strongest instrument for identity protection has the least purchase on the identities that are fabricated.

Assessment. Comprehensive transparency architecture, meaningful platform risk duties, no dedicated liability route for synthetic identity harms, and a discovery burden that rests with regulators and auditors rather than individuals. Strong on paper. Its adequacy depends on whether the Article 50 marking obligations survive contact with open-weight models and metadata stripping, which is an empirical question that the transitional period has not yet answered.

4.3 United States

The United States has no comprehensive federal framework and a rapidly expanding farrago of state law, sectoral federal rules, and constitutional constraint.

At federal level, three developments matter. The Federal Trade Commission's impersonation rule, discussed above, reaches synthetic personas in commercial and governmental impersonation contexts and provides a monetary relief mechanism. The TAKE IT DOWN Act, enacted in 2025, established notice and removal obligations for non-consensual intimate imagery including synthetic imagery, and is notable as the first federal statute to impose a defined takedown timeline on platforms in this domain. Legislative proposals to establish a federal property right in voice and likeness have advanced through committee without enactment as of May 2026, and by their terms would protect real individuals rather than reach fabricated ones.

At state level, the picture is more active and less coherent. Several states have enacted election-related synthetic media statutes requiring disclosure of AI-generated content in political advertising within defined pre-election windows. Others have enacted provisions on synthetic sexual imagery, on AI disclosure in commercial chat interactions, and on the impersonation of licensed professionals. California and Texas have moved furthest, and California's requirement

that certain conversational systems disclose their artificial nature is among the closest existing analogues to a synthetic identity disclosure duty.

The constitutional overlay is the defining feature of the United States position. First Amendment scrutiny has already produced adverse outcomes for election-related synthetic media statutes, with courts finding that provisions imposing platform-level removal duties or restricting political speech containing manipulated media were insufficiently tailored. The doctrine that false statements are not categorically outside constitutional protection means that any prohibition-based approach must be narrowly drawn to a recognised harm, and disclosure mandates must survive scrutiny as compelled speech. This is the central constraint on synthetic identity regulation in the United States and it recommends liability-based approaches, which regulate harm rather than expression, over prohibition-based ones.

Section 230 remains the principal obstacle to platform detection duties. Its protection is directed at liability for third-party content, and the argument that an obligation to verify entity authenticity is distinct from liability for content is available but untested.

Assessment. Fragmentary, constitutionally constrained, and strongest where it operates through consumer protection and civil liability rather than content restriction. The discovery burden rests almost entirely with individuals and with a small number of federal enforcers.

4.4 United Kingdom

The United Kingdom has adopted a sectoral, regulator-led approach and has not legislated a cross-cutting AI statute.

The Online Safety Act 2023 imposes duties on user-to-user services in respect of illegal content and content harmful to children, with a fraudulent advertising duty applying to the largest services. Ofcom's implementation through codes of practice, proceeding in phases through 2025 and 2026, includes provisions relevant to fraud and to inauthentic behaviour, but the Act's architecture is content-centred. A synthetic persona that posts nothing illegal creates no duty, and the fraudulent advertising duty reaches paid placements rather than organic personas.

The Fraud Act 2006 covers synthetic identity deployed for gain, and the false representation offence is drafted broadly enough to reach representations about the existence of the representor. This is doctrinally useful and practically limited by the same attribution problems that constrain enforcement everywhere.

Data protection law, following the reforms enacted in 2025, retains the structural feature noted in relation to the GDPR: it does not reach personas without natural referents.

The government's stated approach continues to favour applying existing regulator powers rather than legislating a new framework, with cross-sector coordination through the central AI function. As of May 2026 this leaves synthetic identity governed by fraud law, sectoral financial regulation, and platform duties that were not designed for it.

Assessment. Principled preference for existing law, resulting in coverage of synthetic identity only where it is instrumental to fraud. No disclosure obligation, no dedicated liability route, and a discovery burden resting with individuals and with sectoral regulators whose remit does not squarely include persona authenticity.

4.5 China

China has the most direct and most prescriptive synthetic media regime in our sample, and it is the clearest example of a disclosure regime with prohibition elements and state-directed enforcement.

The Provisions on the Administration of Deep Synthesis Internet Information Services, in force since January 2023, apply to services that use generative and synthesis technologies to produce text, images, audio, video, and virtual scenes. They require service providers to verify the real identity of users, to obtain consent where a natural person's biometric information is edited, to add conspicuous labels to synthesised content that may cause public confusion, and to maintain records enabling traceability. The Interim Measures for the Management of Generative AI Services, in force since August 2023, extend registration, filing, and content obligations to generative services offered to the public.

The Measures for Labelling AI-Generated Synthetic Content, effective September 2025, are the most significant development for our purposes. They require both explicit labelling, meaning perceptible to users, and implicit labelling, meaning embedded in file metadata, across generated content, and impose obligations on platforms to detect labels, to check for implicit markers where explicit ones are absent, and to label content they assess as probably synthetic. They also prohibit the removal, alteration, or concealment of labels and the provision of tools for that purpose.

This last provision is the single most notable feature in the entire comparative sample. Every disclosure regime faces the circumvention problem identified in Section 5.1. China is the only jurisdiction in our sample that has directly criminalised the circumvention rather than merely mandating the label.

The regime's foundation is real-name verification, which has no analogue in the other jurisdictions surveyed and is not exportable to them. Real-name registration makes synthetic persona creation on domestic platforms materially harder and simultaneously constitutes a comprehensive identification infrastructure with well-documented implications for anonymous speech, political dissent, and minority protection. The regime is effective against synthetic identity partly because it is architecturally incompatible with anonymity.

Assessment. The most complete disclosure and labelling architecture, the only anti-circumvention provision, platform detection duties that are genuine rather than risk-based, and a discovery burden placed squarely on platforms and regulators rather than individuals. It also demonstrates that the most effective available technical answer to synthetic identity, namely verified identity infrastructure, carries costs to civil liberties that the other seven jurisdictions have been unwilling to accept. The lesson for democratic jurisdictions is not to replicate the model but to determine how much of its effect can be obtained without its foundation.

4.6 South Korea

South Korea occupies a middle position, combining a new framework statute with existing election and network provisions.

The Framework Act on the Development of Artificial Intelligence and Establishment of Trust, passed in December 2024 and taking effect in January 2026, is the first comprehensive AI statute in Asia. Its transparency provisions require notification where a service is provided using generative AI and labelling of outputs that are difficult to distinguish from reality, with implementation detail left to presidential decree. The Act combines these duties with a risk-tiering approach and domestic representative requirements for foreign providers.

Korea's election law has for several years prohibited the use of deepfake video, images, and audio in campaign contexts within a defined pre-election window, and this prohibition has been enforced. The Information and Communications Network Act supplies offences relevant to defamation and to false information distributed through information networks, and the Personal Information Protection Act supplies the data protection layer with the same limitation identified elsewhere.

Assessment. A serious framework in early implementation with the substance dependent on subordinate legislation. Transparency obligations reach generative services rather than persona deployment as such. Election prohibitions are among the most operationally enforced in the sample. The discovery burden rests with regulators and, in the election context, with the election commission, which is a meaningful improvement on individual-borne discovery.

4.7 Canada

Canada is the clearest illustration of a legislative lacuna produced by process rather than principle.

The Artificial Intelligence and Data Act, introduced as part of Bill C-27, would have established a federal framework including transparency obligations for systems generating content that could be mistaken for human-created material. The Bill died on the order paper with the prorogation of Parliament in January 2025 and had not been reintroduced in equivalent form as of May 2026. The subsequent creation of a dedicated federal AI portfolio signalled continued policy attention without producing an enacted framework.

In the absence of AIDA, synthetic identity in Canada is governed by the Criminal Code offences of fraud and identity fraud, which carry the structural limitation identified in Section 3.1, by the Competition Act's provisions on false and misleading representations, and by the personal information protection statutes, which again do not reach personas without natural referents. The Canada Elections Act prohibits false statements about candidates in defined circumstances and the impersonation of election officials and candidates, which reaches impersonation of real people rather than fabrication of fictional ones.

Provincial privacy and consumer protection statutes add coverage in specific contexts. Quebec's private sector privacy legislation, following its 2021 reforms, includes automated decision transparency obligations that are among the strongest in North America but do not bear on persona authenticity.

Assessment. No dedicated framework, coverage confined to fraud and misleading representations, and a discovery burden resting entirely with individuals and prosecutors. Canada is a useful case precisely because it shows what the default looks like when comprehensive legislation stalls: the general law applies, and the general law does not fit.

4.8 Brazil

Brazil has advanced the most ambitious rights-based proposal in our sample and has not yet enacted it.

The comprehensive AI bill approved by the Federal Senate in December 2024 and under consideration in the Chamber of Deputies as of May 2026 combines a risk-tiering structure with an explicit rights framework. Its provisions include a right to be informed when interacting with an AI system, a right to explanation, a right to contest automated decisions, and labelling obligations for synthetic content. Notably for our purposes, the proposal adumbrates civil liability provisions for damage caused by AI systems, with allocation of the burden of proof that departs from ordinary negligence principles in favour of affected persons.

That last element is the most directly relevant provision in the entire comparative sample to the recommendation developed in Section 6.2. A liability framework that shifts the evidentiary burden toward the party that deployed the system addresses the discovery problem at its root, because it does not require the harmed individual to prove matters that only the deployer can know.

Pending enactment, synthetic identity in Brazil is governed by the Civil Rights Framework for the Internet, which conditions platform liability on judicial orders in most circumstances, by the General Data Protection Law with its

familiar limitation, by the Consumer Defence Code, which is unusually protective and includes strict liability elements for suppliers, and by electoral regulation. The electoral authority's rules for the 2024 municipal elections prohibited deepfakes in campaign material, required disclosure of AI use in electoral propaganda, and imposed obligations on platforms to remove certain categories of content without judicial order, and this remains among the most assertive electoral synthetic media regime in the democratic world.

Recent constitutional developments narrowing the conditional immunity model in the Civil Rights Framework increase the likelihood that platform obligations in Brazil will become more demanding.

Assessment. The proposal with the strongest theoretical fit to the synthetic identity problem, particularly on liability allocation, and no enacted framework as of May 2026. The electoral regime demonstrates that context-specific prohibition with platform duties is administrable. Discovery burden allocation under the pending bill would be the most favourable to individuals in our sample.

4.9 India

India has proceeded through subordinate instruments and advisories rather than primary legislation.

The Information Technology Rules of 2021, as amended, impose due diligence obligations on intermediaries including obligations relating to impersonation and to content that is patently false. Advisories issued to platforms in 2024 addressed the labelling of synthetically generated material and the risks of deepfakes in the electoral context. Proposed amendments circulated in late 2025 would introduce defined obligations for synthetically generated information, including prominent labelling covering a specified proportion of the display area for visual content, embedded metadata, and user declaration requirements at upload, with significant platforms required to deploy technical measures to verify declarations.

The Digital Personal Data Protection Act 2023, with rules notified in late 2025, establishes the data protection layer and carries the limitation identified throughout. Criminal provisions on cheating by personation, now located in the Bharatiya Nyaya Sanhita, reach impersonation and are subject to the missing-victim problem where the persona is fabricated.

Assessment. A rapidly developing labelling regime driven through executive instrument rather than statute, with unusually specific technical requirements and meaningful platform verification duties in the proposed amendments. The reliance on advisories and rules rather than primary legislation raises questions of legal certainty and of judicial durability. The discovery burden is placed on platforms in the proposed amendments, which is the correct allocation.

4.10 Comparative summary

Jurisdiction	Dominant model	Persona-specific duty	Platform detection duty	Private right of action	Discovery burden
European Union	Disclosure with prohibition elements	Partial (AI Act Art. 50)	Indirect (DSA systemic risk)	Limited	Regulator and auditor
United States	Liability and sectoral prohibition	No	No	Yes, via consumer and tort law	Individual
United Kingdom	Existing law application	No	No	Limited	Individual
China	Disclosure with anti-circumvention	Yes	Yes, direct	No	Platform and regulator
South Korea	Disclosure with election prohibition	Partial	Partial	Limited	Regulator
Canada	None enacted	No	No	Limited	Individual
Brazil	Rights and liability (pending)	Proposed	Proposed	Proposed, with burden shifting	Would shift to deployer
India	Disclosure via subordinate instrument	Proposed	Proposed	No	Platform

Three observations follow from the table.

First, no jurisdiction in the sample has enacted a duty that attaches specifically to the creation or deployment of a persona as such. Every existing obligation attaches either to content or to systems. The persona, which is the unit of harm, is the unit that nothing regulates.

Second, the jurisdictions that place the discovery burden on platforms are China and, prospectively, India and Brazil. Among the jurisdictions with the strongest speech protections, the burden remains with individuals, which is the allocation least likely to produce recourse.

Third, only one jurisdiction has enacted an anti-circumvention provision, and only one has proposed a burden-shifting liability rule. These are the two mechanisms that most directly address the failure modes identified in Section 5, and they exist in isolation from each other in jurisdictions with otherwise divergent approaches.

5. Assessing the Three Models

We assess the strengths and limitations of each model and identify the conditions under which each is most likely to be effective.

5.1 Disclosure-based regimes

Disclosure is the dominant regulatory instrument in this field, and it has become something of a legislative nostrum. There are good reasons for its popularity. It is the least speech-restrictive option available, it places no substantive limit on what may be created, it aligns with established practice in advertising and financial regulation, and it is politically achievable in a way that prohibition is not.

It also carries five weaknesses that, taken together, mean disclosure alone cannot close the recourse deficit.

Circumvention is cheap and the incentive to circumvent is total. A synthetic persona deployed for fraud or influence has no value if it is labelled. The actors most likely to comply with a labelling obligation are those whose use is benign, and the actors causing the harms the rule exists to address are precisely those who will strip metadata, use offshore or open-weight models, or generate outside the regulated perimeter. Provenance metadata survives well within cooperating pipelines and poorly across screenshots, re-encodings, and adversarial removal. Watermarking research has repeatedly demonstrated that robustness against a determined adversary remains unsolved. China's anti-circumvention provision is the only direct legislative response to this in our sample, and its effectiveness has not yet been independently assessed.

Disclosure does not reach persona infrastructure. Labelling regimes attach to outputs. A synthetic identity is an assemblage: a face from one system, text from another, a voice from a third, an account created manually, a history accumulated over months. Labelling each output does not label the entity, and it is the entity that persists and accrues trust.

Notice does not equal comprehension. The evidence from privacy notices, cookie banners, and financial disclosure is consistent and discouraging. Disclosure that is technically present but habitually ignored transfers responsibility to the individual without transferring capability. A machine-readable marker buried in file metadata satisfies a legal obligation and informs nobody, which is disclosure in form and otiose in substance.

Disclosure does not remedy harms already caused. Labelling requirements are prospective and informational. They provide no route to compensation, no mechanism for takedown, and no consequence proportionate to injury. A person defrauded by an unlabelled persona is in the same position after the rule as before it, except that a regulatory breach has occurred alongside their loss.

The liar's dividend. As labelling becomes normalised, the absence of a label acquires evidentiary weight it cannot bear, and the presence of synthetic content in public discourse gives bad-faith actors a general-purpose tool for denying authentic material. Disclosure regimes should be designed with the awareness that they change the epistemic status of unlabelled content as well as labelled content.

Conditions for effectiveness. Disclosure works where the regulated party has a compliance incentive independent of the rule, meaning established firms with reputational and licensing exposure; where the disclosure is perceptible at the point of decision rather than embedded in metadata; where circumvention is itself an offence; and where a remedy attaches to non-disclosure that is proportionate to the harm avoided. Absent those conditions, disclosure produces compliance among the compliant.

5.2 Prohibition-based regimes

Prohibition restricts the creation or deployment of synthetic identities in specified contexts. Its strength is that it does not depend on individual vigilance. Where a category of deployment is prohibited, the question at enforcement is whether the conduct occurred rather than whether an individual was deceived, which lowers the evidentiary threshold considerably.

Prohibition works best where three conditions hold: the context is bounded and identifiable, such as an election period, a regulated financial interaction, or a judicial proceeding; the harm is severe and the legitimate use case is thin; and an institutional actor with monitoring capacity and standing already exists in that context, such as an electoral commission or a financial conduct regulator. The electoral prohibitions in Korea and Brazil satisfy all three and are correspondingly among the more enforced provisions in our sample.

Its limitations are equally clear.

Constitutional exposure. In jurisdictions with strong speech protections, particularly the United States, broad prohibitions on synthetic personas will not survive scrutiny. Recent litigation over election-related synthetic media statutes demonstrates that even narrowly framed provisions face serious tailoring problems, particularly where they impose removal duties on intermediaries or reach political expression.

Definitional overreach and underreach. Prohibitions drafted broadly capture satire, art, research, security testing, and protective anonymity. Prohibitions drafted narrowly are evaded by operations that move one step outside the defined context. The election-period prohibition is evaded by influence operations that build persona credibility for eighteen months and deploy it two days before the window opens.

Enforcement asymmetry. Prohibition is enforceable against domestic, identifiable actors. The majority of high-consequence synthetic identity operations are neither. A prohibition regime with no extraterritorial reach and no platform-facing obligation regulates the least significant portion of the problem.

5.3 Liability-based regimes

Liability assigns responsibility for harms caused by synthetic identities to their creators or deployers. Of the three models it is the only one that operates directly on the economics of the activity, and it is the only one that provides a remedy to a person who has actually been harmed.

Its advantages are substantial. It regulates harm rather than expression, which makes it more durable against constitutional challenge than prohibition. It scales through private enforcement rather than depending entirely on regulatory capacity. It creates incentives for upstream actors to build safeguards, because liability exposure is a cost that firms manage systemically. And it is the only model that addresses the situation of a person who has already suffered loss.

The obstacles are causation, attribution, and access to evidence.

Causation. Synthetic identity harms are frequently diffuse. Where a network of fabricated constituents shifts perceived public opinion, no individual can demonstrate personalised injury traceable to a specific persona. Liability performs well for concentrated harms, such as fraud and harassment of an identified person, and poorly for collective ones.

Attribution. Liability requires a defendant. Operations are structured to prevent identification, and cross-border structuring is the norm rather than the exception.

Evidentiary asymmetry. The facts that establish liability, including which system generated the persona, what safeguards were in place, and who deployed it, are known to the defendant and unknown to the claimant. Ordinary burden allocation therefore renders the claim nugatory even where the substantive rule is favourable. This is the discovery burden expressed in procedural form, and it is the reason the Brazilian proposal's burden-shifting provision matters more than its substantive rules.

Conditions for effectiveness. Liability works where the harmed party can identify a defendant, where the evidentiary burden is allocated to the party with access to the evidence, where the harm is concentrated enough to support individualised proof, and where remedies are available that do not require the claimant to fund protracted litigation. The first and second conditions can be engineered through legal design. The third and fourth require attention to collective and low-value claims, which is why the recommendations in Section 6 include representative action.

5.4 Model comparison

Criterion	Disclosure	Prohibition	Liability
Addresses harms already caused	No	No	Yes
Operates without individual vigilance	No	Yes	Partially
Robust to circumvention	Weak	Moderate	Strong
Constitutional durability	Moderate	Weak	Strong
Reaches foreign actors	Weak	Weak	Moderate, via platform and asset nexus
Addresses collective harms	Partially	Yes	Weak without representative action
Creates upstream design incentives	Weak	Weak	Strong
Depends on regulatory capacity	High	High	Low

The table supports the finding stated at the outset. No single model is sufficient. Disclosure supplies the informational baseline and the evidentiary predicate for everything else. Prohibition handles the bounded contexts where the harm is severe and the legitimate use is negligible. Liability supplies deterrence and remedy. The failure across our sample is not that jurisdictions have chosen the wrong model. It is that they have chosen one model and treated it as complete.

6. Minimum Standards for Adequate Governance

Drawing on the comparative analysis, we propose minimum standards for synthetic identity governance. These are framed as minimum standards rather than as a model law because the jurisdictions surveyed differ too widely in constitutional structure for a single text to be portable. What follows specifies the function each standard must perform and the design features without which it will not perform it.

6.1 Standard One: mandatory disclosure in political, commercial, and journalistic contexts

The rule. Any person who deploys a synthetic identity in a political, commercial, or journalistic context must disclose its synthetic nature in a manner perceptible to the audience at the point of interaction. Removal, alteration, or concealment of a required disclosure, and the provision of tools or services whose principal purpose is to effect such removal, must itself be unlawful.

Why these three contexts. Political contexts because the harm is to collective self-government and no individual claimant will emerge. Commercial contexts because the deception operates on economic decisions and existing consumer protection principles supply a doctrinal foundation. Journalistic contexts because synthetic sources and authors corrupt the mechanism by which societies establish shared facts, and because the harm to the information ecosystem is irreversible in a way that individual harms are not.

Design requirements. Four features distinguish an effective disclosure obligation from a nominal one.

Disclosure must be *perceptible*, meaning visible or audible to an ordinary user at the moment of engagement, not

encoded in metadata alone. Metadata marking should be required in addition, as it enables downstream detection, but it should not satisfy the obligation on its own.

Disclosure must attach to the *persona* rather than to each output. The obligation should run to the account, profile, or persistent identity, so that a single labelled artefact does not launder an unlabelled entity.

Circumvention must be *independently unlawful*. This is the provision that distinguishes China's regime from every other in our sample and it is the provision most worth transplanting, since it can be adopted without any part of the identification infrastructure that surrounds it.

Non-disclosure must carry a *remedy*, developed in Standard Two. A disclosure obligation whose only consequence is a regulatory finding leaves the harmed individual where they started.

Objections addressed. The compelled speech objection is strongest where disclosure is imposed on political expression. It is answerable where the obligation is directed at a factual matter, namely whether the speaker exists, rather than at the content of the speech, and where it is confined to persons deploying personas rather than to platforms hosting them. Disclosure that a communication comes from a non-existent person is closely analogous to established sponsorship identification requirements, which have survived scrutiny for a century.

6.2 Standard Two: civil liability with a rebuttable presumption

The rule. Where a person suffers harm through interaction with a synthetic identity, the creator and the deployer of that identity are liable for the harm, subject to a rebuttable presumption of liability that may be discharged by demonstrating compliance with the disclosure obligations in Standard One and the exercise of reasonable care to prevent foreseeable misuse.

Why a presumption. This is the central recommendation of the paper and it follows directly from the discovery burden analysis. Under ordinary burden allocation, a claimant must establish that the persona was synthetic, identify who created and deployed it, and prove the causal link to their loss. Each of those facts is within the exclusive knowledge of the defendant. A substantive right that requires the claimant to prove facts they cannot access is not a right. Shifting the burden does not make defendants liable for harms they did not cause. It requires the party with the information to produce it.

What the presumption does and does not do. It attaches on proof of two elements: that the claimant suffered legally cognisable harm, and that the interaction from which the harm arose involved a synthetic identity. It does not attach on proof of intent, and it does not require the claimant to establish the identity of the system that generated the persona. It is discharged by evidence of compliant disclosure, which means a fully labelled persona carries no presumptive liability at all, or by evidence of reasonable care proportionate to the deployment context.

Who is a creator and who is a deployer. Creator means the person who generates the persona artefacts or who provides a system whose outputs are used to do so where the provider knew or ought reasonably to have known of the deployment. Deployer means the person who operates the persona in interaction with others. The two are frequently different and both should be reachable, with contribution between them determined by ordinary principles. Providers of general-purpose systems should be reachable only where the reasonable care standard has been breached, meaning where foreseeable persona-generation misuse was not addressed by available safeguards. Nothing in this standard is intended to impose strict liability on model providers for all downstream conduct.

Collective harms and representative action. Because many synthetic identity harms are diffuse, the liability

standard must be paired with a route for collective redress: representative or class action, regulatory action on behalf of affected classes, or a statutory damages floor for defined categories such as election-period deployment. Without one of these, the liability regime will remedy the concentrated harms and leave the systemic ones untouched, which inverts the actual severity ranking.

Interaction with intermediary immunity. The presumption runs against creators and deployers, not hosts. It is compatible with existing intermediary liability frameworks and does not require their amendment. Platform obligations are addressed separately in Standard Three, and deliberately so, because conflating the two has been the principal source of constitutional difficulty in litigation to date.

6.3 Standard Three: platform detection and labelling obligations at scale

The rule. Platforms above a defined user threshold must maintain proportionate systems to detect synthetic identities operating on their services, must label or restrict accounts identified as synthetic, must provide accessible reporting routes and timely response for individuals who believe they have been harmed by a synthetic persona, and must publish auditable metrics on detection, action, and appeal.

Why platforms. Individual enforcement cannot address synthetic identity at scale. This is not a preference for centralised control but an observation about where capability sits. The signals that make synthetic personas detectable, including registration telemetry, device and network fingerprints, behavioural regularities, and the graph structure of persona networks, are visible only in aggregate and only to the operator of the service. No individual, and no regulator without compelled access, can perform this analysis. Placing the discovery burden anywhere other than the platform guarantees that most synthetic identities are never discovered.

Design requirements. Obligations must be *systemic rather than item-by-item*, assessed against the adequacy of the detection system rather than against outcomes in individual cases, which both fits the technical reality and reduces the risk of the obligation operating as a content-removal mandate.

Obligations must be *proportionate to scale and risk*, with a threshold that excludes small services, and calibrated more strictly for services with electoral, financial, or minor-facing exposure.

Obligations must be *auditable*, with independent verification of the metrics published, following the model the Digital Services Act has established for systemic risk.

Obligations must include *due process for the accused*. Detection systems generate false positives, and a wrongly labelled real person suffers a serious dignitary and often economic harm. Notice, human review, and a genuine appeal route are not optional features of this standard. They are what distinguishes it from an arbitrary deplatforming power.

What this standard does not require. It does not require real-name verification. The comparative analysis shows that identity verification is the most technically effective approach available and that its costs to anonymous speech, political dissent, and vulnerable users are not acceptable in a rights-respecting framework. The detection duty proposed here operates on behavioural and provenance signals and is designed to be discharged without knowing who any user actually is. This distinction should be stated expressly in any implementing instrument, because the drift from detection duty to identification mandate is the principal risk this standard carries.

6.4 Standard Four: international coordination mechanisms

The rule. States should establish coordination mechanisms addressing the cross-border dimension of synthetic identity,

comprising at minimum: mutual recognition of provenance and labelling standards, expedited cross-border evidence procedures for attribution, a coordinated approach to jurisdiction over foreign deployers who target domestic populations, and a shared technical baseline for provenance metadata.

Why coordination is not optional here. Synthetic identity operations are inherently cross-border. Generation occurs in one jurisdiction, deployment in another, platform hosting in a third, and harm in a fourth. Every element of the framework proposed above degrades without coordination. Disclosure obligations are evaded by generating outside the perimeter. Liability is evaded by structuring the deployer offshore. Platform duties fracture where provenance standards diverge, because a marker applied under one standard is unreadable under another.

Realistic mechanisms. Comprehensive treaty-making in this field is chimerical on any relevant timescale, and the recommendation is therefore for layered and achievable instruments.

Technical standards first. Provenance and content credential standards are being developed through industry consortia and standards bodies. States should converge on a common baseline and reference it in domestic instruments rather than each specifying incompatible technical requirements. This is the single highest-value coordination step available and it does not require a treaty.

Regulatory networks second. Existing cooperation channels among consumer protection authorities, electoral commissions, data protection authorities, and financial regulators can be extended to synthetic identity without new legal architecture. Enforcement cooperation among like-minded regulators has produced results in adjacent fields and is available now.

Jurisdictional principle third. States should adopt a consistent effects-based principle asserting jurisdiction over deployers who direct synthetic personas at domestic populations, following the model already established in data protection and increasingly in platform regulation. Consistency matters more than reach: divergent assertions produce conflicts of law that sophisticated actors exploit.

Evidence procedures fourth. Attribution is the binding constraint on liability. Existing mutual legal assistance mechanisms operate on timescales incompatible with the lifecycle of a persona network. Expedited procedures for preservation and disclosure of account and provenance data, modelled on those developed for cybercrime cooperation, are a necessary complement to any domestic liability rule.

7. Objections and Responses

“This will chill legitimate anonymous speech.” The framework proposed here is designed around this objection, which is the most serious raised against synthetic identity regulation. Standard Three expressly excludes real-name verification. The definition in Section 2 excludes pseudonymity by requiring the absence of a natural person behind the persona, which means the dissident, the whistleblower, the abuse survivor, and the person writing under a pen name are all outside its scope. The distinction that matters is between concealing which person is speaking and asserting that a person exists where none does. The first is a protected practice. The second is a representation about a matter of fact.

“Detection is technically infeasible, so the platform duty is a duty to do the impossible.” Detection at the level of individual accounts is genuinely difficult and becoming more so. Detection at the level of networks and behaviour is tractable and is already performed by major platforms, as their own transparency reporting on coordinated inauthentic behaviour demonstrates. The standard proposed is systemic adequacy rather than per-account accuracy,

which is precisely the framing that accommodates the technical reality. A duty to maintain a proportionate detection system is not a duty to achieve perfect detection.

“Compliance costs will entrench incumbents.” This is a genuine concern and it is why thresholds and proportionality are built into Standard Three. It is also worth noting that the liability standard, unlike the platform standard, applies without a size threshold, because a small deployer causing serious harm should not be immune. The asymmetry is deliberate: obligations that require infrastructure are scaled, and obligations that require honesty are not.

“Rebuttable presumptions of liability are an extraordinary departure.” They are established in product liability, in employment discrimination, in certain environmental regimes, and in consumer protection across several of the jurisdictions surveyed. The common feature of those contexts is evidentiary asymmetry, meaning that the defendant holds the relevant information. Synthetic identity presents that asymmetry in an unusually acute form. The presumption is discharged by disclosure compliance, which means a well-behaved deployer faces no meaningful exposure.

“Definitional line-drawing will capture art, satire, and research.” The definition requires deployment in a context where the audience would reasonably assume engagement with a natural person, which excludes fiction and declared virtual personas by construction. Security research, red-teaming, and academic study should be protected by express exemption with reasonable conditions. Any instrument that omits such an exemption will impede the research community whose work makes detection possible in the first place.

“Synthetic personas have legitimate protective uses.” They do. Investigative journalism, law enforcement operations, safety research, and the protection of at-risk individuals all involve fabricated personas. Each of these is an authorised or professionally governed activity and each should be handled by express exemption. The existence of legitimate uses is an argument for careful carve-outs rather than against regulation, and the same argument was made and answered in the development of every impersonation offence now in force.

8. Implementation Pathways

The standards above are not equally difficult to enact, and sequencing matters.

Immediately available. Provenance standard adoption requires no primary legislation and can proceed through procurement requirements, regulatory guidance, and existing standards processes. Regulatory cooperation among consumer, electoral, and financial authorities can be extended under existing mandates. Anti-circumvention provisions can be attached to disclosure obligations already enacted or in transitional application, and represent the highest-value amendment available to jurisdictions that have already legislated on labelling.

Achievable within a legislative cycle. Context-bounded disclosure obligations for political, commercial, and journalistic deployment have precedent in every jurisdiction surveyed and face the least constitutional difficulty when framed around the existence of the speaker rather than the content of the speech. Platform detection duties can be added to existing systemic risk frameworks where those exist, which is the shortest available path in the European Union and the United Kingdom.

Requiring sustained effort. The liability standard is the most consequential and the most demanding. It requires primary legislation, careful drafting of the presumption and its rebuttal, provision for collective redress, and coordination with existing intermediary immunity frameworks. Brazil’s pending bill is the closest existing model and its progress through the Chamber of Deputies should be watched closely by drafters elsewhere.

Sequencing note. Enacting disclosure without liability produces the failure mode identified in Section 5.1, in which obligations bind the compliant and the harmed remain without remedy. Enacting liability without disclosure leaves claimants with a right and no evidentiary predicate. The two should be enacted together or, where that is impossible, disclosure should be enacted with an explicit legislative commitment to the liability layer and with anti-circumvention provisions included from the outset.

9. Policy Recommendations

Recommendation 1. Define synthetic identity as a distinct regulatory object. Legislatures should adopt a definition that turns on the absence of a natural referent, machine-generated constitution, persistence, and contextual assumption of humanity, and should expressly exclude pseudonymity, declared virtual personas, fiction and satire, and authorised protective use. Regulating personas as content or as systems has left the unit of harm unregulated in every jurisdiction surveyed.

Recommendation 2. Attach disclosure obligations to personas, not only to outputs. Where labelling regimes exist or are in transitional application, the obligation should be extended to persistent identities, so that an account, profile, or persona presented as a natural person carries the disclosure rather than each artefact it produces.

Recommendation 3. Criminalise circumvention of provenance and disclosure mechanisms. The removal, alteration, or concealment of required labels and embedded provenance markers, and the provision of tools or services whose principal purpose is to effect such removal, should be independently unlawful. This is the single highest-value amendment available to jurisdictions that have already legislated on labelling, and it can be adopted without any element of identity verification infrastructure.

Recommendation 4. Establish civil liability with a rebuttable presumption against creators and deployers. The presumption should attach on proof of cognisable harm arising from interaction with a synthetic identity, and should be discharged by evidence of compliant disclosure or the exercise of reasonable care proportionate to the deployment context. This addresses the evidentiary asymmetry that currently makes substantive rights unenforceable.

Recommendation 5. Pair liability with collective redress. Representative action, regulatory action on behalf of affected classes, or statutory damages for defined categories such as election-period deployment should accompany the liability standard. Without collective redress, the regime will remedy concentrated harms and leave the systemic harms, which are the more severe, without any route at all.

Recommendation 6. Impose proportionate, auditable platform detection duties with due process guarantees. Services above a defined threshold should maintain and publish auditable detection systems for synthetic identities, with notice, human review, and appeal for accounts that are labelled or restricted. Duties should be assessed on systemic adequacy rather than per-account accuracy, and instruments should state expressly that the duty is not to be discharged through identity verification.

Recommendation 7. Converge on a common provenance standard. States should adopt a shared technical baseline for provenance metadata and reference it in domestic instruments rather than specifying divergent requirements. This is achievable without treaty-making, is the highest-value coordination step currently available, and is a precondition for every other cross-border mechanism proposed here.

Recommendation 8. Establish expedited cross-border attribution procedures. Existing mutual legal assistance

timescales are incompatible with the lifecycle of persona networks. States should establish expedited preservation and disclosure procedures for account and provenance data, modelled on cybercrime cooperation instruments, together with a consistent effects-based jurisdictional principle over deployers who direct synthetic personas at domestic populations.

10. Conclusion

The synthetic identity problem has an unusual shape. It is not a case of a technology outpacing the law in the ordinary sense, where the rules exist but the facts have changed. It is a case where the rules were built on an assumption that the technology has removed. Identity law assumes that behind every persona there is a person, and that assumption held for long enough that it was never written down. Generative systems have falsified it, and in doing so they have exposed how much of the architecture of identity protection was resting on something nobody thought to state.

The consequence is a governance vacuum with a particular character. It is not that synthetic identity harms are unregulated in the sense of being permitted. It is that they are unreachable. A person defrauded by a persona that does not exist has suffered a real loss, has no route to a defendant, and frequently does not know that the loss was caused by deception at all. The recourse deficit here is close to total, and it is total because the discovery burden sits with the party least capable of bearing it.

Closing that gap requires three things that no jurisdiction in our sample has yet assembled together. It requires disclosure obligations that attach to personas and that make circumvention itself unlawful, because a labelling regime without an anti-circumvention provision regulates only those who would have complied anyway. It requires liability with a shifted evidentiary burden, because a right that depends on proving facts held exclusively by the defendant is not a right. And it requires platform-level detection, because the signals that reveal synthetic personas exist only in aggregate and only where the aggregate is visible.

Each of these mechanisms exists somewhere. Anti-circumvention provisions exist in China. Burden-shifting liability is proposed in Brazil. Platform-level systemic duties exist in the European Union. Election-context prohibitions are enforced in Korea and Brazil. What does not exist anywhere is the combination, and the combination is what the problem requires, because each mechanism fails in the way the others are designed to prevent.

The window for assembling it is not indefinite. Synthetic identity harms compound through a mechanism that other technology harms do not share: each successful deployment degrades the baseline assumption that makes quotidian interaction possible. Once populations adjust to the possibility that any counterparty may be fabricated, the cost is borne not by the perpetrators but by every real person who now has to prove they exist. That is the harm that has no plaintiff and no remedy, and it is the one this framework is ultimately intended to prevent.

Annex A: Comparative Coding Framework

Each jurisdiction was coded against six questions. The framework is offered here so that it may be applied to jurisdictions outside our sample.

Q1. What conduct is regulated? Coded as: content generation, system provision, persona deployment, or none. This question isolates the unit of regulation and is the source of the paper's first comparative finding, namely that no

surveyed jurisdiction regulates persona deployment as such.

Q2. Who bears the obligation? Coded as: provider, deployer, platform, or user. Multiple codes permitted. Regimes that place obligations only on providers are vulnerable to open-weight and offshore generation.

Q3. What triggers the obligation? Coded as: generation, publication, interaction, or harm. Trigger points determine whether an obligation attaches before or after the harm occurs and therefore whether the regime is preventive or remedial.

Q4. What remedy attaches? Coded as: administrative penalty, criminal sanction, civil damages, injunctive relief, takedown, or none. Regimes with administrative penalties only provide no route to individual recourse.

Q5. Who may invoke the remedy? Coded as: regulator, prosecutor, affected individual, representative body, or none. This question determines whether enforcement capacity is finite.

Q6. Where does the discovery burden sit? Coded as: individual, regulator, platform, or shifted to defendant. This is the determinative question for whether the regime can close the recourse deficit, and it is the question that existing comparative literature most consistently omits.

A regime that scores “persona deployment” on Q1, “deployer and platform” on Q2, “interaction” on Q3, “civil damages and injunctive relief” on Q4, “affected individual and representative body” on Q5, and “platform or shifted” on Q6 would satisfy the minimum standards proposed in Section 6. No regime in our sample scores this profile.

Annex B: Model Statutory Language

The following is offered as drafting reference rather than as model law, and requires adaptation to the constitutional and doctrinal context of any adopting jurisdiction.

B.1 Definition

“Synthetic identity” means a persona presented as a natural person where (a) no natural person corresponds to the persona; (b) the persona is constituted wholly or substantially by artefacts generated by an artificial intelligence system, in any modality or combination of modalities; (c) the persona persists across time or across multiple interactions; and (d) the persona is deployed in a context in which a reasonable person would assume that they are engaging with a natural person.

A persona is not a synthetic identity by reason only that (i) a natural person uses a name other than their legal name; (ii) the persona is expressly and prominently identified as artificial at the point of interaction; (iii) the persona appears in a work of fiction, satire, art, or commentary in a context where a reasonable person would not assume engagement with a natural person; or (iv) the persona is deployed in the course of authorised law enforcement activity, professionally governed investigative journalism, bona fide security research, or the protection of a person at risk of harm, subject to such conditions as may be prescribed.

B.2 Disclosure obligation

1. A person who deploys a synthetic identity in a political, commercial, or journalistic context shall disclose that the identity is synthetic.
2. A disclosure under subsection (1) shall be (a) perceptible to an ordinary recipient at the point of interaction; (b)

associated with the persona and not solely with individual items of content; and (c) accompanied by machine-readable provenance information conforming to the prescribed technical standard.

3. Machine-readable provenance information alone does not satisfy subsection (1).

B.3 Anti-circumvention

1. No person shall remove, alter, obscure, or conceal a disclosure or provenance marker required under section B.2.
2. No person shall manufacture, supply, offer, or make available any device, product, service, or tool the principal purpose of which is to contravene subsection (1).
3. Subsection (1) does not apply to acts undertaken in the course of bona fide security research or forensic analysis.

B.4 Rebuttable presumption of liability

1. Where a person suffers harm arising from an interaction with a synthetic identity, the creator and the deployer of that synthetic identity are liable for that harm.
2. Liability under subsection (1) is presumed on proof that (a) the claimant suffered harm of a kind recognised by law; and (b) the interaction from which the harm arose involved a synthetic identity.
3. The presumption in subsection (2) is rebutted where the defendant establishes that (a) the requirements of section B.2 were complied with in respect of the synthetic identity; or (b) the defendant exercised reasonable care, proportionate to the context of deployment and to the foreseeability of the harm, to prevent the use of the synthetic identity in a manner giving rise to harm of that kind.
4. A provider of a general-purpose artificial intelligence system is a creator for the purposes of subsection (1) only where the provider knew, or ought reasonably to have known, that outputs of the system were being used to constitute synthetic identities, and failed to exercise reasonable care to prevent that use.
5. Nothing in this section affects the operation of any enactment limiting the liability of an intermediary service provider in respect of content provided by a recipient of the service.

B.5 Platform detection duty

1. A designated service shall establish and maintain systems and processes, proportionate to its size and to the risk profile of its service, to identify synthetic identities operating on the service.
2. A designated service shall (a) label or restrict accounts it identifies as synthetic identities; (b) provide an accessible mechanism by which a person may report a suspected synthetic identity and receive a substantive response within a prescribed period; (c) provide notice, human review, and an effective right of appeal to any account labelled or restricted under paragraph (a); and (d) publish, and submit to independent audit, prescribed metrics relating to detection, action, and appeal.
3. Compliance with subsection (1) shall be assessed by reference to the adequacy of the systems and processes maintained, and not by reference to the outcome in any individual case.
4. Nothing in this section requires or authorises a designated service to verify the legal identity of any user.

Sources and Further Reading

This paper was prepared from primary sources, meaning statutory text, implementing regulations, published regulatory guidance, and judicial decisions, with secondary commentary used for interpretive context only. Positions are stated as at May 2026. Given the pace of legislative and judicial development in this field, readers should verify current status before relying on any jurisdictional description. Full citations for the instruments referenced are available on request from the Institute.

European Union. Regulation (EU) 2024/1689 (Artificial Intelligence Act), in particular Article 50 and associated recitals; Regulation (EU) 2022/2065 (Digital Services Act), in particular the systemic risk provisions and associated Commission guidance; Regulation (EU) 2016/679 (General Data Protection Regulation); Directive 2005/29/EC (Unfair Commercial Practices Directive), in particular Annex I.

United States. Federal Trade Commission Rule on Impersonation of Government and Businesses, 16 CFR Part 461, and the associated supplemental notice of proposed rulemaking; Federal Trade Commission Rule on the Use of Consumer Reviews and Testimonials, 16 CFR Part 465; the TAKE IT DOWN Act; 18 U.S.C. 1028A and associated case law on the “another person” element; 47 U.S.C. 230; California and Texas election, chatbot disclosure, and synthetic imagery statutes; recent federal district and appellate decisions on First Amendment challenges to state synthetic media laws.

United Kingdom. Online Safety Act 2023 and Ofcom implementation codes of practice; Fraud Act 2006, section 2; Data Protection Act 2018 as amended; government policy statements on the cross-sector AI regulatory approach.

China. Provisions on the Administration of Deep Synthesis Internet Information Services (2023); Interim Measures for the Management of Generative Artificial Intelligence Services (2023); Measures for Labelling AI-Generated Synthetic Content (2025) and the associated mandatory national standard; Cybersecurity Law real-name provisions.

South Korea. Framework Act on the Development of Artificial Intelligence and Establishment of Trust (2024, effective 2026) and implementing decrees; Public Official Election Act deepfake provisions; Act on Promotion of Information and Communications Network Utilization and Information Protection; Personal Information Protection Act.

Canada. Bill C-27 and the Artificial Intelligence and Data Act as introduced and amended in committee; Criminal Code sections on fraud and identity fraud; Competition Act deceptive marketing provisions; Personal Information Protection and Electronic Documents Act; Canada Elections Act; Quebec Act respecting the protection of personal information in the private sector, as amended.

Brazil. PL 2338/2023 as approved by the Federal Senate; Lei 12.965/2014 (Marco Civil da Internet) and recent constitutional jurisprudence on intermediary liability; Lei 13.709/2018 (LGPD); Código de Defesa do Consumidor; Tribunal Superior Eleitoral resolutions on artificial intelligence in electoral propaganda.

India. Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 and subsequent amendments and draft amendments concerning synthetically generated information; Ministry of Electronics and Information Technology advisories on synthetic media; Digital Personal Data Protection Act 2023 and rules; Bharatiya Nyaya Sanhita provisions on cheating by personation.

Technical and interdisciplinary. Content provenance and authenticity standards developed through industry consortia and international standards bodies; the research literature on watermarking robustness and adversarial removal; platform transparency reporting on coordinated inauthentic behaviour; security research on synthetic persona use in credential fraud and remote employment schemes.

The Digital Dignity Institute of California is an independent research institute based in San Francisco and engaged globally. It is independent of government and commercial interests. Its work spans global convenings, original research, policy advice, and structural consulting, and is directed at developing frameworks that centre human beings and make artificial intelligence answerable to them.

Correspondence: connect@digitaldignityinstitute.com

© 2026 The Digital Dignity Institute of California. This paper may be reproduced with attribution.