Identity Assurance in an Era of Digital Disruption: Planning a Controlled Transition

HARVARD Kennedy School
MOSSAVAR-RAHMANI CENTER
for Business and Government

Identity Assurance in an Era of Digital Disruption: Planning a Controlled Transition

John Fiske
Harvard Kennedy School

June 2023

M-RCBG Associate Working Paper Series | No. 205

Mossavar-Rahmani Center for Business & Government
Weil Hall | Harvard Kennedy School | www.hks.harvard.edu/mrcbg

HARVARD Kennedy School
MOSSAVAR-RAHMANI CENTER
for Business and Government

Identity Assurance in an Era of Digital Disruption:

Planning a Controlled Transition

John Fiske

June 2023

About the author: John Fiske is a Senior Fellow at the Harvard Kennedy School Mossavar-Rahmani Center for Business and Government. He is also employed by Meta as Director of Data Protection. His work as a Senior Fellow is independent from his work at Meta. Conversely, he in no way represents Meta in this effort.

Special thanks to Satwik Mishra and Hugh Grant-Chapman, both of the Harvard Kennedy School, for their invaluable research assistance on this project, to John Haigh for his mentorship, and to the Senior Fellow cohort of 2022-2023 for their thoughtful input.

Table of Contents

Executive Summary

If you do not know history, you think short term.
If you know history, you think medium and long term.

-Lee Kuan Yew

This paper endeavors to make the case that we should begin a controlled transition towards widespread online identity assurance over the next decade. While pressure building in this direction over the past five years, the democratization of generative AI tools has altered the risk-benefit balance of the 'anonymous' Internet. AI-powered services will soon perfectly replicate online human speech, text, appearance, and activity so that they are indistinguishable from humans. These capabilities will potentially empower a wave of harms such as misinformation and disinformation, manipulation of elections, security hacks, fraud, hate speech, shaming, blackmail, etc. These harms will undermine trust and damage the online social and economic fabric if not mitigated.

Thus, this paper asserts that, whether driven by government decree or private sector self-protection, individuals and enterprises will soon need to provide greater levels of identity assurance to interact online. Note that identity assurance is not the same thing as ID verification. Identity assurance is *'the ability of one party to determine, with varying levels of assurance, that an identity claim made by another party can be trusted'*. The 'claim' can be any characteristic, such as "I am a human" or "I am above 18 years old" or "I am a student at University XYZ". Establishing a trusted framework for identity assurance can greatly help mitigate online harms, though it also raises issues around privacy, surveillance, equity and competition fairness.

The current catalyst for change is that many observers are calling for regulators to require bots to self-identify. This is indeed an important assurance mechanism - online citizens should know when they are interacting with a bot, the rules of engagement with that bot, and on whose behalf the bot is operating. However, such laws are necessary but not sufficient. Necessary, because they will define acceptable societal boundaries and establish legal accountability, but insufficient for two reasons:

1. Laws alone will not adequately prevent bot misrepresentation because good actors will follow the rules, bad actors will not - they will misrepresent their bots as human and take advantage of the user trust created by good actors. Given the democratization of AI tools (where powerful AI models can be run on laptops from basements around the world) bad actors can easily move beyond the jurisdiction of effective regulators.
2. Identity assurance is a larger topic than just bots. Bots act on behalf of people, businesses or governments, and it is just as important that we verify whom they are serving as the fact that they are bots. This need, plus a host of other pressures suggest that we need a comprehensive framework that allows bots, people (in both personal and professional capacities), businesses, media sources and governments to establish appropriate facts about themselves and the entities with whom they are interacting.

What is to be done? Conceptually, we need to flip the assurance paradigm. Instead of assuming that people are who they say they are (and making users or service providers figure out the reality) we need to reverse the assurance paradigm and assume the worst: *we should assume that all entities with whom we interact are in fact untrustworthy bots.* By assuming the worst, and forcing entities to positively prove basic identity claims, we can protect online integrity and safety.

Of course, for this assurance model to be in any way practical requires that people have an easy, private and secure way to prove their personhood, age or any other claim they make about themselves. A few years ago, this would have been an insurmountable challenge, but as of 2023 we see a viable technical framework materializing over the next decade.

Two trends are converging to make this possible.

1. Countries around the world are embracing digital identity credentials, typically issued by governments but sometimes offered via federated arrangements with banks or other trusted institutions.
2. There have been significant technological advances to offer privacy-protective identity assurance at scale. Global standards are emerging, such as ISO18013-7 (online presentation) and the W3C Verifiable Credentials standard, which support different levels of government control. The NIST 800-63 guidelines describe security and trust standards, and the new eIDAS2 defines identity assurance standards for Europe.

Together, these two trends will likely make global, privacy-protective identity assurance a viable possibility over the next decade. While there are many potential technical, legal and operational models which governments and the industry might adopt, the most privacy-protective approaches will utilize identity assurance standards to give users maximum control over the use and sharing of their credentials. They all offer users some form of ID wallet (giving them power to share identity information outside of government control or visibility), secure management of their information, and support of selective disclosure (which lets them share only the minimal amount of information needed).

However, even if, optimistically, we can move toward a privacy-protective identity assurance framework, there are still a number of societal risks we should be aware of and proactively mitigate. We analyze seven categories of risk:

1. Government Surveillance – the risks that governments might (legally or surreptitiously) surveil the online activity of their citizens
2. Inequity – the risks that some governments might not provide some of their citizens with adequate identity credentials for them to operate online
3. Data Protection Violations – the risks that online service providers might misuse identity information they collect
4. 'Metaverse' Risks – the emerging risks from AI-supported environments including bot misrepresentation, identity asymmetry, failed recall, confusing cross-credentialling, etc.
5. Inconsistent Standards – the risks that users can be misled or confused by a range of different identity assurance protocols, poor useability, or varying authentication requirements
6. Inadequate Support – the risks that users might not be supported adequately in a multi-stakeholder ecosystem, especially in cases of identity theft or user death

7. Untrustworthy Ecosystem – the risk the entire ecosystem is deemed untrustworthy by users, governments or online service providers, and the trust framework breaks down

How might we begin this transition? Given the speed with which the digital environment is changing, we suggest that a public-private partnership would be most effective. Governments alone tend to move slowly, and corporations tend to act in their narrow self-interest. But together, they can combine the policy goals of societal protection with the execution capabilities of industry to implement an identity assurance framework quickly. A public-private framework would likely include an oversight body to monitor and lead the effort, an industry code of conduct to establish guidelines for fair use and user support, as well as technical and legal standards to harmonize user protections.

Aligning disparate large technology firms, online service providers and national governments toward a common goal is admittedly a complex challenge. But the issue of identity assurance is upon us and the urgency of Internet safety in the age of AI is growing. With or without any plan, we will likely need greater levels of identification. If governments do nothing, identity assurance may become a market segment dominated by big tech firms and susceptible to national surveillance, abuse and technical lock-in. As Internet safety is in interests of all people, companies and governments, leading stakeholders (especially in the US, EU and India) should align on basic principles of oversight, data protection and technical interoperability in order plan a controlled transition. We should act soon before it is too late.

I. Identity Assurance in Five Concepts

What is 'Identity'?

If someone asked you to 'identify yourself', what would you do? What information would you provide? In the United States, we commonly give our name, date of birth, address, telephone number and present a government-issued ID as proof. In India, one might provide a government-issued Aadhar number and submit a fingerprint to prove it. Both of these mechanisms are effective – you are making a claim about your identity and providing credentials to verify that claim. The credentials uniquely identify you amongst all other people, they are persistent – i.e., they don't easily change - and they provide a high level of confidence (at least to the degree to which we trust the government issuing the credentials).

In contrast, if someone asked you to 'describe your identity', what would you say? You might start with our name and where you live but identity is much larger than that. You might talk about our family relationships, age, nationality, ethnicity, religion, favorite sports team, financial situation, race, sexuality, community membership, schooling, friend group or hobbies. In short, 'identity' can encompass any of innumerable characteristics that describe various facets of our being.

In this paper we take the term 'identity' to include both ideas:

- Identifying information – the set of information by which a person can uniquely identify themselves (or be uniquely identified by others). This might include government issued identity numbers, biometric signatures (fingerprint, iris or facial recognition) and genetic information.
- Personal characteristics – the much larger set of characteristics that describe various aspects of a person and their life. These characteristics are not unique to any individual - although it is quite possible to uniquely identify someone by combining numerous personal characteristics about them.

What is Identity Assurance?

In plain language, identity assurance offers users and businesses a mechanism to confirm facts about anyone they interact with online. But 'identity' is a complex concept, and the terminology is complicated. In this article, we define identity assurance as *'the ability of one online entity to determine, with varying levels of assurance, that an identity claim made by another online entity can be trusted'*. An "online entity" can be a person, organization, or bot, each representing either themselves or another entity.

Note that an "identity claim" is NOT the same as identity verification – rather it is a claim to any characteristic, such as "I am a human" or "I am above 18 years old" or "I am a student at University XYZ". "Identity verification" (also known as "identity proofing") means the verification of a (usually government-issued) credential which uniquely identifies the holder. Also note that "identity assurance" is different from the following concepts:

- "Authentication" is the process of confirming that a person is the same person who created an account, or otherwise registered with a service (but does not imply that the user made any claims about their real-world identity).
- "Digital identification" is a broad term that can mean anything from issuing a credential in digital form, to the credential itself, to the process of establishing or verifying one identity.
- "Identity presentation" refers to how a user presents themselves to other people inside of a service. Depending on the specific terms of service, the user may remain anonymous to other users, use a pseudonym, or be required to share their verified real name – but will usually need some form of identity assurance with the service provider.

Levels of Assurance

Just as in real life, online identity information conveys varying degrees of trust. Given the sophistication of identity theft today, it is hard to ever be 100% certain about anyone's identity online so we must ascribe varying levels of confidence to a claim based on the credential itself, the user, or other extenuating circumstances. A government issued ID verifying a name and address might give us high confidence (which we might require in order to issue a loan or sell someone a car for example). A library card plus an email address might give us medium confidence (enough to let them use our free Internet service, but little more). Someone's claim that they graduated from a university might only warrant low trust, but an alumni email address with the university domain might push that up to medium confidence. If we already know an individual well and they have a track record of telling us the truth, or we have other information about them which is consistent with their new claims, then confidence levels might be increased. We call this confidence level a 'Level of Assurance'.[1]

Identity Claims and Trust Triangles

When people or businesses state certain facts about themselves this is called 'making an identity claim'. Examples might be 'a person claiming to be a licensed doctor', or 'the mother of a certain child', or to be 'a person named John Doe'. Claims without proof have a low level of assurance, so we need a mechanism to verify a claim. In practice today there are many ways to do this, but recent identity assurance standards have embraced a three-party system which was elegantly labelled the 'trust triangle' by proponents of one architecture[2]. The three parties are the credential 'Issuer' (typically a trusted institution such as a government agency, a bank, a university, etc.), the 'Holder' (i.e., the person who owns the credential and is trying to verify a claim) and the 'Verifier' (the person or business or government that needs to verify a claim). The Holder will make a claim to the Verifier supported by a credential given by the Issuer. Beyond that conceptual triangle there is a large variation in how things

[1] The NIST Digital Identity Guidelines 800-63-3 expands the concept of Levels of Assurance into three related concepts (Identity Assurance Level, Authentication Assurance Level and Federation Assurance Level) to give US federal agencies more flexibility in designing identity solutions. Other standards also have more nuanced definitions of Level of Assurance, and in practice, these variations are helpful; but for reasons of simplicity, this paper uses 'Level of Assurance' as a general concept to convey the confidence level of a given identity claim.

[2] The Self-Sovereign ID model is a relatively recent architecture that relies on cryptography and distributed ledgers to verify identity claims privately and is reflected in W3C and the eIDAS standards. The same three-parties are supported in ISO standards as well. https://academy.affinidi.com/what-is-the-trust-triangle-9a9caf36b321

actually work. Some standards require the Verifier to proof a credential with the Issuer; other standards embrace a 'verified credential' model whereby the credential can be validated anonymously via encrypted tokens. These technical options are beyond scope for this paper.

Selective Disclosure

Selective disclosure is the idea that credentials and trust systems should be designed to give Holders (aka users) the ability to select what information they disclose. For example, when asked to prove that they are over age 21, a person should not be required to show their full driver's license with name address, and a host of other personal identifying information; ideally, they should have a mechanism to prove they are over age 21 and disclose no other information. Similarly, users will need the ability to prove that they are human (i.e., their "personhood"), their nationality, their approximate location, without revealing other details to the Verifier. Selective disclosure is one of the primary protections against corporate overcollection of data and government surveillance.

II. An Era of Digital Disruption: 2023-2030

Emerging Needs for Stronger Identity Assurance

Change is coming to our online world. All signs suggest the 2020s will be a decade of profound disruption between the rise[3][4] of AI and AI-powered services[5], the global embrace of digital currencies, the mainstreaming of augmented and virtual reality, the imminent robotics revolution and the advent of quantum computing. This wave of technological development will inevitably change both how we live our personal digital lives, and how the broader economy operates[6].

As we visualize this near future, one recurring theme is that digital identity is a foundational challenge for many of the most impactful developments. To describe a few of them:

- *Personal digital assistants* will serve humans with ever-more intimate and important services.[7] They will likely use the latest generative AI technology and be indistinguishable from humans in appearance, speech, or gesture. Moreover, they will have the power to act as an agent on behalf of people in both the digital and the real world and ultimately perhaps the ability to filter (or distort) information flows reaching an individual. For them to achieve widespread acceptance requires a trustworthy framework to identify them as digital assistants, to define the limits of their agency, to clarify who has built the technology and who is operating it, and to establish clearly who they are serving (and thus who is accountable for their actions).
- *Synthetic media* will leverage generative AI (visual and text) to create and distribute information. This naturally creates deep concerns about misinformation[8] (more on that later) but these services will likely generate value for specific audiences. The author and publisher of information needs to be established, in some cases, to help platforms and individuals determine whether the information is trustworthy (and who to hold accountable for abuse).

---

[3] "Since the advent of Deep Learning in the early 2010s, the scaling of training compute has accelerated, doubling approximately every 6 months." Available at: Sevilla, J. *et al.* (2022) "Compute Trends across three eras of machine learning," *2022 International Joint Conference on Neural Networks (IJCNN)* [Preprint]. Available at: https://doi.org/10.1109/ijcnn55064.2022.9891914.

[4] "The amount of compute used in the largest AI training runs has been increasing exponentially with a 3.4-month doubling time (by comparison, Moore's Law had a 2-year doubling period). Since 2012, this metric has grown by more than 300,000x (a 2-year doubling period would yield only a 7x increase)." Available at: *Ai and Compute AI and compute*. Available at: https://openai.com/research/ai-and-compute (Accessed: April 19, 2023).

[5] The private investment in AI in 2021 totaled around $93.5 billion) Zhang, D., Lynch, S. and Miller, K. (2022)) Zhang, D., Lynch, S. and Miller, K. (2022) *AI index 2022*, *Stanford Institute for Human-Centered Artificial Intelligence*. Available at: https://hai.stanford.edu/research/ai-index-2022 (Accessed: April 19, 2023).

[6] The World Economic Forum estimates that 70% of new value created over the next decade will be based on digitally enabled platform business models." World economic forum, Strategic Intelligence. Available at: https://intelligence.weforum.org/topics/a1Gb0000001SH21EAG (Accessed: April 19, 2023).

[7] The global market for Intelligent Virtual Assistants estimated at US$7.3 Billion in the year 2022, is projected to reach a revised size of US$47.4 Billion by 2030, growing at a CAGR of 26.4% over the analysis period 2022-2030. "https://www.researchandmarkets.com/reports/5030189/intelligent-virtual-assistants-global#tag-pos-6 ltd, R. and M. (no date) *Intelligent Virtual assistants - global strategic business report*, *Research and Markets - Market Research Reports - Welcome*. Available at: https://www.researchandmarkets.com/reports/5030189/intelligent-virtual-assistants-global#tag-pos-6 (Accessed: April 19, 2023).

[8] Disinformation Researchers Raise Alarms About A.I. Chatbots - The New York Times (nytimes.com)

- *Professional bots* will emerge to help businesses operate, sell and service their customers. Like personal digital assistants, they will be perfectly human-like, and may be used to directly represent individuals in a commercial capacity or a larger business. An identity framework should identify them as bots, define the limits of their agency and who they are serving, whether an individual or a business.

- *Virtual reality* technologies will create immersive digital experiences, with high-fidelity connections between people or bots, each representing themselves through avatars or filters. Multi-party "metaverse"[9] environments will give agency to businesses and people to create, communicate, deliver services, and exchange money directly with each other, largely outside of the oversight or control of any single service provider. Depending on the rules of the environment, it may be impossible to tell who is human vs bot, and a clear identity framework is essential to prevent bad actors from misrepresenting themselves or other people.

- *Physical robots* will likely move beyond today's menial tasks and begin to operate more autonomously in the world of humans. In their next phase, they may need to do things like schedule a repair, order supplies, or reserve other assets or services. Beyond that, we may soon have kitchen robots which do our shopping, cooking and cleaning. For any of these tasks we again need a clear identity framework to identify them as robots, define the limits of their agency and establish ownership and accountability for their actions.

- *Digital currencies and digital assets* are poised to proliferate globally. This is a large topic, but identity issues are central to many services and capabilities:

    - New transaction models with digital currencies (e.g., bitcoin, digital yuan, etc.) will require some form of identity authentication to comply with Know Your Customer / Anti-Money Laundering (KYC/AML) laws. Taxation methods may also adapt to digital currency transactions with some countries requiring ID verification.

    - Traditional digital payments also require identification, and the growth of digital ID is expected in markets "where mobile-first services help citizens access banks, loans, insurance, and government services."[10]

    - Microtransactions may soon be possible, creating new business models, but requiring very low-cost identity confirmation mechanisms.

    - New legal structures (e.g., DAOs – Decentralized Autonomous Organizations) are already in operation and require identification of the principals and owners of an asset.

    - The exchange of digital assets (non-fungible tokens, digital skins, art, etc.) requires new ownership rules, especially considering the transfer of these assets across borders. This again requires a trusted identity assurance framework.

- *Digital Social Structures.* Over time, digital society may evolve social structures similar to the real-world such as eVoting, eJustice, and virtual statehood. Some countries have already begun these developments. Estonia offers e-citizenship and eVoting.[11] China has launched a pilot test

[9] By the term 'metaverse' we do not mean to imply any particular corporate vision of the 'metaverse'

[10] What does the future of digital identity look like? - Raconteur

[11] ID-card - e-Estonia

for 'AI judges'[12] in some courts. Some Swiss cantons offer eVoting.[13] Russian expatriate groups are exploring developing a "virtual Russian state-in-exile."[14] All of these developments require identification with high levels of assurance.

Chains of Accountability

One other thread connecting these emerging needs is that the nature of identity itself may need to evolve. In addition to containing trusted information about a person or a business itself, 'identity' may eventually encompass information about their relationships. Just as a social graph maps the connections between different stakeholders, identity might extend to relationships for which the individual is accountable. For example:

- Parent-child relationships – establishing guardianship over a minor
- Human-bot relationships – ensuring accountability for bot activity
- Business-employee relationships – allowing an employee to identify as representing a business
- Author-work relationship – allowing confirmation of authorship to improve information integrity
- Owner-asset relationship – establishing ownership of assets (especially digital assets)
- Etc.

It is out of scope to develop this concept further, but we need to be open to reimagining the credential itself as we move into the next decade.

Ongoing Pressures Toward Identity Assurance

So, identity assurance is an emerging priority as we look ahead. Looking backwards, it has also been growing in importance. The Internet's anonymity has been liberating for many but it has also given rise to numerous online harms. These harms have been steadily driving greater levels of online identity assurance:

- *Current online social harms*: Fraud[15], mis-/disinformation, anonymous abuse, hate speech, bullying, identity theft/misrepresentation, blackmail, spambots, and related threats are rampant online today[16]. Virtually all online businesses are wrestling with these challenges, and are under heavy pressure to manage these problems more systematically. The anonymity of users is a structural challenge when trying to prevent these harms.

- *Youth protections*: Youth safety online is also a top issue for regulators and online platforms[17]. Again, a critical obstacle to age-appropriate safeguards is the lack of an easy and precise way to verify user ages. While there are useful ways to estimate user ages, they have significant ranges of error at the key age thresholds. As children are unlikely to be issued credentials any time

[12] China Now Has AI-Powered Judges — RADII
[13] E-voting - online voting in Switzerland
https://www.economist.com/international/2022/08/09/much-of-russias-intellectual-elite-has-fled-the-country
[15] The US FTC reported that online fraud had impacted 2.4 million Americans in 2022, resulting in $8.8B in losses (a 30% increase over 2021). 41% of the cases involved impersonation or identity theft.
[16] Ofcom Online Nation 2022 reports that 45% of online users in the UK encounter potentially harmful 'contact harms' (such as trolling, unwanted sexual messages and bullying, abuse or threats) in a given four week period
[17] The European Commission launched a new 'Better Internet for Kids' initiative in 2022

soon, an additional safeguard would be an ability to confirm the age of older users and then restrict unidentified user access to youth services by default.

- *Regulatory pressure*: There are numerous pending regulations (e.g., the Digital Services Act in Europe, the EU AI Act, the UK Online Safety Bill, the California Age-Appropriate Design Code Act, etc.) which will hold service providers accountable for online harms to users. This in turn will pressure them to confirm certain aspects of user identities, both to create more effective safety controls and to hold bad actors accountable.

As a few examples of this trend, Elon Musk stated his intention to authenticate all users to eradicate spambots when announcing his intent to purchase Twitter[18]. In 2019, Mark Zuckerberg stated his ambition to authenticate 1 billion Facebook users in order to combat misinformation (a goal that has since been abandoned for privacy and other reasons, though Meta has recently launched a paid verification service[19]) . The US Combating Organized Retail Crime Act of 2022[20] would require all online sellers of a certain size to verify their addresses. Meanwhile, the Metaverse Standards Forum has made Identity one of its five priority workstreams[21].

The Need for a New Paradigm

The democratization of AI changes everything. If the industry was moving slowly towards identity assurance in the past five years, the democratization of generative AI has turbocharged progress. With the open-source release of large language models (LLMs) and image generation models, individuals can tune and configure their own AI models. Combined with increasing edge bandwidth,[22] cheaper storage, and greater computing power, smaller AI models can be run on laptops from anywhere in the world. This will allow bad actors to convincingly misrepresent their own identity for purposes of fraud, manipulation or theft. It will also empower them appropriate the identity of other people (or businesses or governments) in order to spread misinformation, cause reputational damage, defraud, etc. Because democratization enables geographic fluidity, bad actors can now easily move beyond the oversight of national authorities or enforcers.

This is expected to have three systemic effects on the flow of misinformation in particular:

1. Increased volume, as AIs can mass produce content very easily, and personalize it ad infinitum
2. Increased speed, as projects which used to take hours can now be done by generative AIs in seconds
3. Increased fidelity, as the most powerful AIs can now generate text and video which is indistinguishable from human outputs

---

[18] Fung, B. (2022) *Elon Musk wants to 'authenticate all real humans' on Twitter. here's what that could mean | CNN business*, *CNN*. Cable News Network. Available at: https://www.cnn.com/2022/04/28/tech/elon-musk-authenticate-all-real-humans/index.html

[19] Roth, E. (2023) *Facebook and Instagram are testing selling you blue checks for $12 a month*, *The Verge*. The Verge. Available at: https://www.theverge.com/2023/2/19/23606268/meta-instagram-facebook-test-paid-verification (Accessed: April 19, 2023)

[20] *H.R.9177 - 117th Congress (2021-2022): Combating organized retail crime.* Available at: https://www.congress.gov/bill/117th-congress/house-bill/9177 (Accessed: April 19, 2023).

[21] Morrison, R. (2022) *How will digital identity work in the metaverse? Tech Monitor*. Available at: https://techmonitor.ai/focus/how-will-digital-identity-work-in-the-metaverse (Accessed: April 19, 2023).

[22] Nielsen's Law of Internet Bandwidth (nngroup.com) suggests that a high-end user's connection speed grows by 50% per year.

The insufficiency of regulation. Recognizing these risks, many observers in early 2023 are calling for regulation requiring bots to self-identify[23][24]. This is indeed an important assurance mechanism - online citizens should know when they are interacting with a bot, the rules of engagement with that bot, and on whose behalf the bot is operating. However, laws are necessary but not sufficient. Necessary because they will define acceptable societal boundaries and legal accountability, but insufficient for three reasons:

1. Laws alone will not adequately prevent bot misrepresentation because good actors will follow the rules, but bad actors will not - they will misrepresent their bots as human and take advantage of the trust created by good actors.
2. Identity assurance is a larger topic than just bots. Bots act on behalf of people, businesses or governments, and it is just as important to verify whom they are serving as the fact that they are bots.
3. As discussed below, a real solution requires more than just regulation. It will require government leadership and industry partnership to develop meaningful safeguards.

Flipping the assurance paradigm. Instead of focusing exclusively on regulation, we need to think about assurance more broadly. What combination of laws, technical standards, oversight bodies and industry codes of conduct can most effectively provide users and business safe and private identity assurance?

As a first step to framing that solution, we need to conceptually flip the assurance paradigm. Instead of legally mandating bots to self-identify and assuming compliance, we need to reverse the assurance paradigm and conservatively assume the worst: *we should assume that all entities with whom we interact are in fact untrustworthy bots.* This mirrors the security models embraced by online banking and cybersecurity businesses, which through low- or zero-trust models have greatly improved corporate security and safety. By assuming the worst, and forcing entities to positively prove their identity claims, we will have much higher levels of assurance. Paradoxically, if implemented well, this framework can improve personal privacy protections by giving users power over their own identity information, securing the information exchange, and setting guidelines for the use of that information.

Of course, for this assurance model to be in any way practical requires that people have an easy, private and secure way to prove their personhood, age or any other claim they make about themselves. A few years ago, this would have been an insurmountable challenge, but as of 2023 we see a viable technical framework materializing and becoming operational over the next decade. Two trends are converging to make this possible.

Parallel Developments

Firstly, the global transition to digital identity credentials is allowing citizens to authenticate their identities online more easily and securely, reducing transactional friction.

[23] https://crsreports.congress.gov/product/pdf/IF/IF11333 The US Congressional Research service note that online platforms need to "expand the means of labelling or authenticating content" to combat deepfakes. (Accessed: May 25, 2023)
[24] In a recent study, subjects could not tell the difference between fake Twitter accounts and real ones and tended to think the AI accounts were less likely to be fake than the genuine ones. The authors state "We need new methods to deal with this [such as] ID verification and other safeguards" https://studyfinds.org/bots-social-media-fake-accounts/ (Accessed: May 25, 2023)

- *Global government embrace of eID*: Today, more than 50 countries have initiatives underway to develop or issue digital national credentials, and user acceptance of digital credentials and wallets is increasing rapidly. Juniper Research estimates[25] digital identity document users will expand from 4.2 billion in 2022 to 6.5 billion in 2026, with much of the growth coming from developing countries in Africa and Asia. Many government initiatives are driven by the efficiencies of e-government and require their citizens to have a digital credential to access those services. In the US and Canada, 33 states are issuing or in the process of trialing[26] mobile drivers' licenses (mDL). Europe is establishing an EU-wide standard (eIDAS 2)[27] which will ensure technical interoperability and legal recognition for national digital IDs in 2024. India's massive Aadhaar identity system recently surpassed 1.3 billion users.[28]
- *Private sector issuance*: Private sector partners are also rolling out digital credentials. In some countries, private companies issue national digital credentials directly under the oversight of the government (in Nordic banks, for instance).[29] Companies are also acting independently, as Mastercard, for example, is establishing its own global identification service.[30] Many private companies offer digital "membership credentials" that can be treated as identity credentials for appropriately low levels of assurance.

Secondly, new technology is available offering privacy-protective identity assurance at scale. A full review of technical options is beyond scope for this paper, but advances in encryption have enabled a new class of authentication technologies to support various trust models.

- *Digital wallets*: From a user perspective, the most visible development is availability of a "digital ID wallet" which makes heavy use of encrypted tokens and/or public/private key encryption to enable secure and private authentication. Wallets give users:
    - Maximum possible control over the sharing and authentication of credentials – they choose with whom they share identity information.
    - "Selective disclosure" allows users to verify certain facts about themselves (e.g., that they are human, or that they are above a certain age) without having to share a full credential.
    - Credential flexibility, as almost any type of identity claim can be supported (for example a familial relationship with another person, a claim of membership in a group, ownership of a bot, etc.).
    - Simpler transactions, as wallets can potentially eliminate the need for most username-password authentication steps. [31]
    - Improved user security via device-level security (2- or 3-factor authentication) and encrypted communications.

[25] Expect digital ID to balloon in users and revenue by 2026: Juniper Research | Biometric Update
[26] Implementation Tracker Map - mDL Connection
[27] *Eidas regulation Shaping Europe's digital future*. Available at: https://digital-strategy.ec.europa.eu/en/policies/eidas-regulation (Accessed: April 19, 2023).
[28] Aadhaar Dashboard (uidai.gov.in)
[29] Sweden BankID: what it is and how it works - Wise
[30] Digital Identity Services | ID Network (idservice.com)
[31] Different studies suggest the average American Internet user has between 100 and 150 online accounts requiring a password

    - Minimal reliance on government verification for most transactions (reducing surveillance risk).

- *ID authentication standards*: While wallets are the most visible identity assurance technology, standards-based infrastructure is required to support end-to-end secure ID assurance. Trust technology has been maturing for the past decade and several new technical standards will be finalized in the 2023 – 2025 timeframe. Leading standards, such as ISO 18013-5 (mDL)[32] and 18013-7 (online presentation)[33] and the W3C Verifiable Credentials standard (closely related to the 'Self-Sovereign ID framework)[34] reflect different levels of centralization by which governments can issue and authenticate credentials. The NIST 800-63 guidelines[35] describe security and trust standards for the US government, and the eIDAS2 standard[36] will define infrastructure and wallets requirements for Europe for decades to come. These models provide governments with greater choice regarding interoperable technical solutions and facilitate stronger privacy protections for users.

These trends suggest new possibilities to rethink our approach to identity assurance. We'll explore the potential societal impacts of this sort of change in a moment – but first let's address one common concern.

But… What About Anonymity?

Some people have a visceral (and understandable) reaction to the idea of identity assurance, concerned that this might mean the end of anonymity on the Internet. This a complex topic but let's attempt to untangle it from identity assurance in a few summary points:

- In practice, today's Internet is only vaguely anonymous – user activity is tracked and logged by numerous technologies on the devices, networks and services we use. Most of this data is not directly tied to a named individual, but it easily can be, so it is hard to claim that we currently have an 'anonymous Internet'. For the foreseeable future, users who desire anonymous Internet access will still need to use technical protections such as a Tor browser, VPNs, cookie blockers etc. to go online without being tracked.
- Identity Assurance is essentially a mechanism for users and service providers to make and verify online identity claims. Privacy-protective incarnations allow users to confirm any claim from the least personal (e.g., mere personhood) to the most uniquely identifying (e.g., full identity verification). So, assuming countries and users embrace privacy protective features, the rules about how platforms use this tool are critical to the question of anonymity. If rules require minimal identification, then anonymity is likely to be strengthened – i.e., a simple claim of 'personhood' or 'nationality' is far less identifiable than current ID sharing practices. So in short, the availability of a privacy-protective identity assurance mechanism does not weaken

[32] ISO/IEC 18013-5:2021 - Personal identification — ISO-compliant driving license — Part 5: Mobile driving license (mDL) application
[33] *ISO/IEC AWI TS 18013-7* (2021) *ISO*. Available at: https://www.iso.org/standard/82772.html
[34] Verifiable Credentials Data Model v1.1 (w3.org)
[35] NIST Special Publication 800-63 Digital Identity Guidelines | NIST
[36] European digital identity (eID): Council makes headway towards EU digital wallet, a paradigm shift for digital identity in Europe - Consilium (europa.eu)

anonymity, but the industry needs guidelines to ensure that the tool is not misused, or else anonymity will suffer.

- Lastly, in the view of the author, we should proactively strengthen anonymity safeguards in order to protect free speech and free information access. A number of suggestions for strengthening the technical and legal protections of anonymity are outlined in the Risk section below. But we should also be clear about anonymity's boundaries, and the fact that truly anonymous users (who refuse any level of identity assurance) will be low trust. Some likely implications of this are that:
    - Only a small percentage of Internet usage will be completely anonymous (i.e., refusing to validate personhood)
    - Services offered to completely anonymous users will likely be a limited subset of the services available with more identity assurance.
    - Trust in anonymous authors will be low and this will curtail their ability to publish widely (unless perhaps they have a trusted party, such as a major publisher, vouching for them).

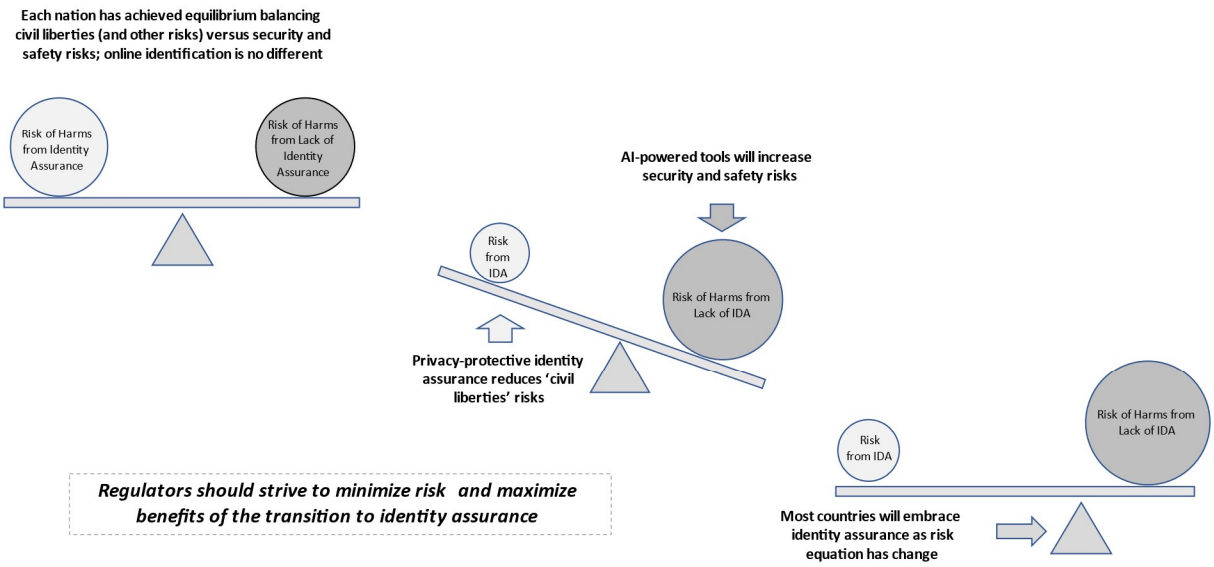
Balancing Risks and Benefits

The question of anonymity raises broader questions about the societal risks and benefits of Identity Assurance. These are well-known tensions - Privacy vs surveillance. Civil liberty vs security. Anonymity vs accountability – which have been debated for decades. Each country has precariously found an equilibrium suitable for their own socio-political values. As we consider rebalancing identity assurance online, we should recognize that each nations' values will differ and thus their appetite for various risks will also be different. That said, and as we described above, the decade ahead promises three structural changes which will alter the risk/benefit balance for all countries:

1. The risks stemming from *not having* identity assurance will increase significantly as ever more potent harms threaten all stakeholders - users, governments, and businesses.
2. The risks of *having* identity assurance will likely decrease significantly as technological advances increase user privacy and reduce the risk of surveillance and abuse.
3. The utility of identity assurance will likely rise as users adopt wallets for identity assurance, online authentication, user rights, and financial transactions.

These three changes will not resolve the civil liberties issues noted above but will change the associated risk levels. Here is a simplified graphic to reinforce that idea.

Increased AI Risks and Reduced Privacy Risks of Identity Assurance are Driving Transition to a New Equilibrium



If implemented fairly, identity assurance can help individuals navigate their digital lives safely, confident in the knowledge of who they are interacting with, and in control of their personal identification. It could:

- Reduce online harms, enabling online service providers to better combat malicious behavior;
- Enable greater trust, as individuals will have greater certainty about the people, institutions and digital services with whom they are interacting.
- Improve efficiency, as the same technology can be used for both virtual and real-world use cases, and standardized technology and processes will improve ease of use for users.
- Increase user privacy protections, security and ability to control their own person information

However, if implemented poorly, identity assurance can increase other dangers. For example, it might:

- Enable government surveillance and political repression by enabling governments to track user online activity[37].
- It might empower corporate exploitation if there are no rules restricting what they can ask of users and what they do with the data provided

[37] *5 problems with national ID cards*. *American Civil Liberties Union*. Available at: https://www.aclu.org/other/5-problems-national-id-cards (Accessed: April 19, 2023).

- It might expand global inequity as people without identity credentials are excluded from online life

The question then becomes: *how can engineer a solution to minimize risk of increased identity assurance while maximizing the benefits online safety, privacy, and freedom of expression?*

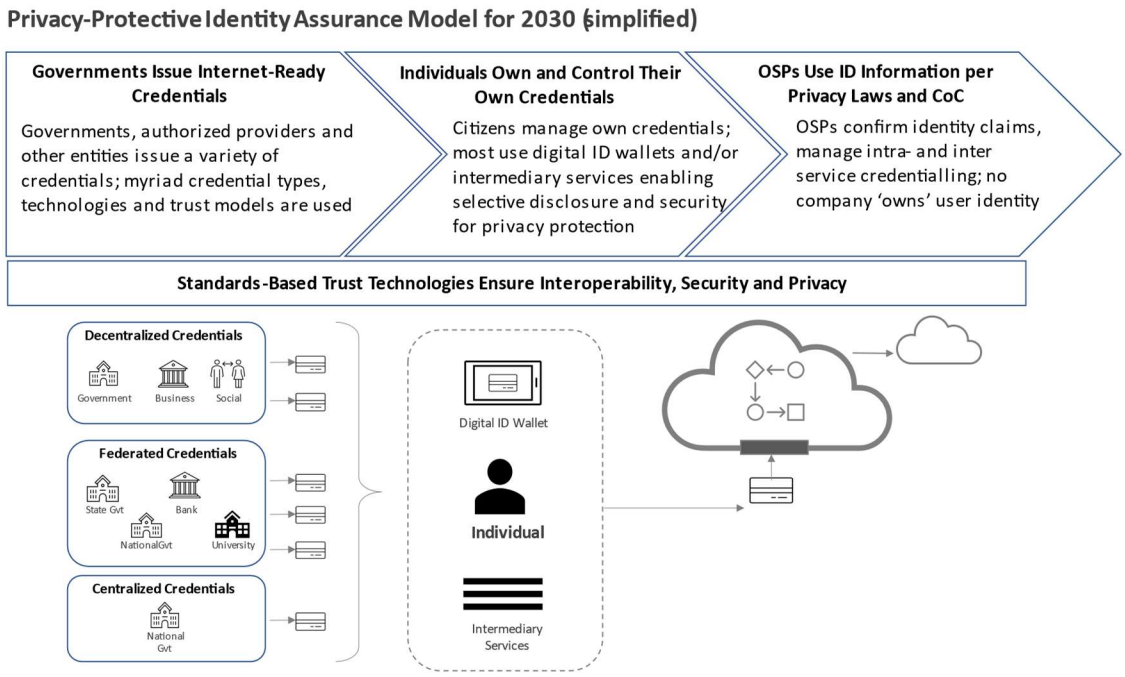
In Section III we look at the key characteristics of a privacy-protective identity assurance model entails; Section IV explores the various societal risks which should considered, as well as possible mitigations; and in Section V we propose steps to start the transition in a controlled manner.

III. Designing Privacy-Protective Identity Assurance

User-Centric Identity Assurance

It is hard to predict exactly what identity assurance technology will look like in the year 2030. The technology is still developing very fast, there are various competing technical standards, and there are disparate visions for the role of government. But that said, the outlines of a privacy-protective identity assurance model are increasingly clear.

The graphic below depicts a greatly simplified picture of what we might achieve by 2030. Note that this is just one possible outcome (and an admittedly optimistic one) but this directional vision has been validated as "likely" by stakeholders from technology firms, regulators and civil society leaders. Some countries (notably China) have chosen a different path whereby identity assurance, in practice, means ID verification for all online activity, and becomes the cornerstone of an online surveillance function. In China's case, this is the basis of their social credit scoring system.[38] But for countries who wish to protect their citizens' personal freedom and privacy, identity assurance can be a critical safeguard.



[38] Data from the Chinese central bank showed that the system has grown to cover 1.1 billion individuals and over 60 million enterprises and organizations. While Chinese government researchers at the end of 2020 complained that local efforts to build social credit systems were "disjointed and inconsistent ", the government also announced that in 2019 it blocked the country's untrustworthy from purchasing plane and high-speed rail tickets more than 30 million times. From "Surveillance State: Inside China's Quest to Launch a New Era of Social Control"

At this altitude, the picture is of a 'framework of frameworks' – i.e., there will be many viable methods to achieve the same goal of privacy-protective identity assurance. The important characteristics to note include:

- *Heterogeneity.* Each nation will inevitably choose their own path and timeframe regarding issuing digital credentials to their citizens. Some will digitize rapidly; others will issue paper credentials throughout the decade. In some countries, ID issuance will remain a purely governmental responsibility; in others, the private sector will play an important role. Some trust models will be highly centralized (e.g., India's Aadhar) some will be federated (e.g., the EU's eIDAS2 model) and others fully decentralized (e.g., Bhutan recently announced a decentralized 'self-sovereign' national ID platform). So we should expect (and embrace) a heterogenous set of trust models and credential issuance frameworks. But we should strongly encourage governments to issue credentials suitable for online identity assurance.
- *Standards-based technology.* There will likely be a similarly diverse set of identity assurance standards (e.g., ISO18015-7, eIDAS2, W3C, etc.) including some which do not exist yet today. An individual's choice of standard may be driven by government decree, market availability, price, technology requirements or their own preferences; regardless, online service providers and businesses should be expected to support a diverse set of technologies.
- *User Control.* Users in this model will (ideally) be given control over the use and sharing of their credentials, rights to access, delete or change their credentials, end-to-end security throughout the identity assurance process and most importantly the power of selective disclosure, which enables them to share only the minimum data required for a certain level of assurance.
- *Norms for Identity Assurance.* Lastly, digital businesses will need established norms concerning identity information. We expect they will use identity assurance in many ways to support their customers and protect themselves but need to use the data responsibly. Appropriate use can be defined by regulation, but voluntary codes of conduct would likely be more effective because of the lack of globally harmonized regulation and oversight, and the slow pace of regulatory development.

Responsibilities of Key Stakeholders

The success of the model rests on four groups of stakeholders fulfilling interdependent responsibilities.

Governments

National governments are responsible, first and foremost, for issuing their citizens "Internet-ready" credentials by which we mean credentials that can be used to support standards-based online identity assurance. Ideally these would be issued in digital format, but even paper or plastic IDs can be made "Internet-ready" through a trust framework which makes them useable in standards-based digital wallets. This trust framework can be run by state governments or by other authorized organizations. It also might be made available to citizens of a repressive regime from an outside entity (likely a qualified international NGO or civil society organization) which could create alternate channels for those citizens to authenticate their IDs outside of the view of their own government.

Governments will each need to pursue their own path to digital identification, so we expect a range of models (centralized, federated, and decentralized) will yield a large variety of credential types, issuers and trust models. Whichever model is used, governments should enshrine user protections in privacy law, and embrace secure, privacy-protective technical standards. Some examples:

1. Even without wallets, India's Aadhaar system is supporting baseline user privacy protections. Aadhar lets users generate a Virtual ID, a 16-digit random number mapped to the Aadhaar number on a temporary basis. Users can provide that 16-digit number (instead of their Aadhaar number) to relying parties, significantly strengthening their online privacy.[39] Though Aadhar relies on a centralized database (and is thus potentially vulnerable to hacking) it stores only minimal user data and is protected by laws[40] prohibiting the national government from tracking authentication activity. Eventually, users may have the option to use an Aadhaar ID in a digital wallet, giving them even greater protection, but this is still a few years away.
2. Several US states and other countries are embracing the ISO18013-5 digital wallets (provided currently by Apple, Google, and other wallet providers)[41]. Governments issue mobile drivers' licenses (mDLs) to their citizens who store them in ID wallets on their phones and can use them in person in lieu of a physical license. The ISO 18013-7 standard[42] (expected in 2023) will make it possible to present an mDL to a verifier over the internet to the websites, supporting "unattended use cases" such as identity proofing, authentication, attribute presentation, and single sign on. This can facilitate online use and reduce the current dependence on passwords and other authentication technologies.
3. The EU is finalizing the eIDAS 2 digital wallet[43] standard (largely based on the W3C standard for verifiable credentials[44]). This, coupled with new EU-wide eIDAS 2 credential standards, could become a global standard, as it promises to support a federated trust structure (with great flexibility) and strong privacy protections.
4. Many other countries, from Bhutan[45] to Ethiopia[46], are designing, testing, deploying and improving digital formats of their national and state credentials.

Whatever the mechanism, governments need to act quickly to offer privacy-protective credentials to their citizens. Governments that lack resources might rely on international NGOs for support. As the Secure Identity Alliance noted in a recent whitepaper: "Governments are becoming increasingly responsible for assuring the proper governance of their national identity digital ecosystems. This entails structuring appropriate trust anchors, considering the legal liability landscape, and adopting international standards for greater interoperability."[47]

---

[39] https://uidai.gov.in/resources/uidai-documents/circulars,-notifications-office-memorandums. html

[40] Section 32(3) of the Aadhaar Act 2016 specifically prohibits UIDAI from "controlling, collecting, keeping or maintaining any information about the purpose of authentication either by itself or through any entity."

[41] Kelts, D. (2022) *Successful adoption of mobile ID hinges largely on protection of Citizen Privacy*, *Successful adoption of mobile ID hinges largely on protection of citizen privacy*. International Association of Privacy Professionals. Available at: https://iapp.org/news/a/successful-adoption-of-mobile-id-hinges-largely-on-protection-of-citizen-privacy/ (Accessed: April 19, 2023).

[42] *ISO/IEC AWI TS 18013-7* (2021) *ISO*. Available at: https://www.iso.org/standard/82772.html (Accessed: April 19, 2023).

[43] *Eidas regulation Shaping Europe's digital future*. Available at: https://digital-strategy.ec.europa.eu/en/policies/eidas-regulation (Accessed: April 19, 2023).

[44] *Verifiable credentials data model V1.1 W3C*. Available at: https://www.w3.org/TR/vc-data-model/ (Accessed: April 19, 2023).

[45] https://www.biometricupdate.com/202302/bhutan-launches-self-sovereign-biometric-digital-id-crown-prince-first-to-enroll

[46] https://shega.co/post/parliament-approves-digital-id-law/

[47] On the Road to User-Centricity: Digital Identity in the Electronic Wallet Era (secureidentityalliance.org)

Users

Users must ultimately be accountable for their own identities – they will make the decision whether to share credential information with another entity. For them to make those decisions, they will need basic technical education and trustworthy information about the relying party. Ideally, the steps and interfaces to ID assurance will become standardized and universal, with clear "signposting" to indicate the integrity of the ecosystem and ensure fair bidirectionality. Users also need clear support frameworks (outlining who is accountable to help users in the event of identity theft, device loss, death, etc.), which will be discussed later.

Standards-Based Trust Technologies

At the time of writing, there are two major standards for identity assurance (ISO and W3C) and others are emerging. Over time, a few standards will likely predominate, and leading wallets will presumably support multiple standards. One current challenge is identity assurance from more than one device. Today, digital ID wallets only run on smartphones, so ID assurance on other devices requires a work-around. Also, issuers usually issue only one instance of a credential (e.g., mDLs) so users are again forced to use their phones. Eventually the technology will need to seamlessly extend to other devices for both convenience and equity reasons.

It is also conceivable that intermediary trust services may fulfill the role of the digital wallet, and related concepts, via a cloud-based service. Of course, these services would need to be highly trustworthy, but for people who require unusual levels of support, redundancy, security, or other personal services, this might be viable. Service providers might also offer a service like "ID escrow" where they could securely store a user's credentials, and only share those credentials in the case of certain events occurring or a smart contract executing.

Lastly, there are visionary concepts being developed, such as the Trust Over IP Framework,[48] which envisions a decentralized trust layer across the entire Internet. This would potentially bring broad benefits like support for multiple ID assurance standards, password-less authentication, private digital connections, and verifiable origins for digital goods, information, and payments. If this trust layer does materialize, it could simultaneously simplify and improve the privacy and the security of the Internet and enable a host of new trusted services to emerge. However, the path to widespread deployment of this framework is uncertain, so in the 2030 timeframe we consider it an outlier scenario.

Online Service Providers

Lastly in the value chain, *service providers are accountable to treat identifying information responsibly.* In some regions (especially the EU) there are strong laws (and active enforcement) to define the responsibilities for service providers to process identifying data. These compliance requirements should be globalized as much as possible, either by lawmakers in unprotected countries developing new policies, or by the service providers committing to a binding code of conduct.

Service providers have responsibility to manage identity information in three phases of their service:

[48] The ToIP Technology Architecture Specification - Trust Over IP

1. *Identity collection and use:* In the initial ID assurance process, companies should only collect the minimum data required for a given level of assurance, have clear retention and deletion practices, establish policies for acceptable use/purpose limitation and support user rights of access, deletion, restriction of processing, etc.

2. *Intra-service cross-credentialling:* service providers should develop and abide by rules concerning how they authenticate a user to various internal services and features, or across multiple user personas.

3. *Inter-platform identity transfers:* When exchanging information with other platforms (between 'metaverses' and/or with the transfer of digital goods for example) fair transparency guidelines should be developed and followed.

Given the liability that comes with identity credentials, most large platforms are cautious with credentials. For instance, Meta deletes identity data after a year and provides an option for users to ask the platform to do this in a month. It also complies with its own data minimization and purpose limitation policies, which are similar to GDPR requirements.[49] However due to the lack of global standards, each service provider creates their own institutional approach. An industry code of conduct would help ensure that all entrants are held to the same level of responsibility.

Challenges with the Current Assurance Model

Many service providers are already verifying ID information for large numbers of users, mostly for account reset and support purposes, and to enable financial transactions.[50] A reasonable question then is: *Why won't the current ID authentication model meet emerging needs?*

In short, the current model is functional but not ideal.

A service provider's Terms of Services (ToS) define the levels of authentication required, which vary from a simple email confirmation to full identity verification. Users establishing credit accounts likely need full ID verification; users on a dating app may need their name verified, and a "selfie pic"; users who have been hacked or who have forgotten their passwords may need to verify their names and addresses. Per the ToS, service providers give users a list of acceptable credentials (unique for each country) including passports, drivers licenses, utility bills, or even library cards.[51] The user typically takes a photo of the credential and emails or uploads it to the authentication provider, who verifies it to the appropriate level of certainty. Sometimes users' credentials are verified with the issuer or cross-checked against a list of restricted individuals. Each company will set its own policies for acceptable verification, pursuant to local laws, and typically delete the credential once it has been confirmed to minimize security and privacy risks.

Service providers often outsource this authentication work to specialist third-party providers, and this model is currently functioning reliably at scale. Millions of users around the world have their IDs

[49] https://about.fb.com/wp-content/uploads/2022/07/Privacy-Within-Metas-Integrity-Systems.pdf
[50] ID authentication is required for Know-Your Customer (KYC) and Anti-Money Laundering (AML) legal compliance.
[51] For a recent report on common credentials used for data subject access requests, see: https://home.bigid.com/hubfs/Whitepapers%20and%20Data%20Sheets/State-of-Data-Rights_BigID_IAPP_Report.pdf?hsLang=en

authenticated every day, and there are relatively few mishaps. However, there are several shortcomings:

1. *Privacy risks:* Without selective disclosure, users are obliged to over-share their credentials, often providing a complete photo ID (with address, biometric details, photo, etc.) just to verify their name or age range. In countries without privacy laws, users act on faith that the company is handling their credentials responsibly.

2. *User burden:* This process is burdensome to users who must find appropriate credentials, convey them to the company, engage with the support teams as needed, and so forth. This can take anywhere from a few minutes to several hours, depending on the complexity of the authentication. Moreover, users must repeat this process each time they need to authenticate.

3. *Cost:* This process is expensive. Depending on many factors (service types, volume, levels of authentication, credential-types, languages, etc.), the cost per transaction for a US user can currently range from $20-$50 per transaction.[52] The effects of this are:

    a. Service providers today minimize the numbers of users they authenticate.

    b. Cost is a barrier to entry for smaller companies, raising anti-competition concerns.

    c. Authentication may only make sense economically to support high-value customers, so it creates a disincentive to serve poor regions.

4. *Lack of industry guidelines:* Interestingly, the principles of privacy law (and antecedents such as the 1970s OSHA standards) were developed in a pre-Internet world[53]. At that time, it was assumed that banks, hospitals, businesses and governments had already confirmed the identity of their customers during the course of regular business, so lawmakers were more focused on how those organizations managed information after collection. Thus, the issue of identification itself is barely considered in standards such as the GDPR but treated as a minor topic within the "processing of personal data." Even today, in the draft American Data Privacy and Protection (ADPP) Act, credential information is simply treated as one more type of "sensitive category data"[54] and is given no broader guidance regarding the foundational topic of identification.
5. *Inequity:* There are roughly 1 billion individuals around the world without government-issued IDs, often women and low-income populations of developing countries,[55] but also disadvantaged residents of developed countries. Lacking credentials, these individuals also lack access to important services (basic financial services, government services, etc.).

Given the likelihood that users will need to confirm elements of their identity much more frequently in the near future, this decade is critical to move to the next generation of ID assurance.

[52] Human-supported authentication is estimated to cost $20-50 per person fully loaded. Real time interviews may cost significantly more. Note: the author has not verified these estimates which were made by authentication experts.
[53] *The Job Safety Law of 1970: Its passage was perilous* (no date) *DOL*. Available at: https://www.dol.gov/general/aboutdol/history/osha (Accessed: April 19, 2023).
[54] https://www.congress.gov/bill/117th-congress/house-bill/8152/text#H4D5FF218B1204DB79D51CD3E9996A096
[55] The global identification challenge: Who are the 1 billion people without proof of identity? (worldbank.org)

IV. Identity Assurance Risks (And Possible Mitigations)

In Section III we looked at the key characteristics of a privacy-protective identity assurance model. In this section we will analyze the various societal risks which should be considered, as well as possible mitigations. Designing a system with appropriate protections is complicated by several factors:

- Numerous different standards are emerging.
- Myriad solutions are coming onto market from vendors serving each stage of the value chain.
- Divided global governance (i.e., no regulator has global jurisdiction), so each country is figuring out its own approach, often at odds with others.
- Minimal alignment between leading online platforms and technology companies on potential rules or standards.

These points suggest that this transition will likely be messy but by considering the risks ahead of time and implementing key mitigations, some harms may be minimized.

This section reviews seven major categories of risk:

I. Government surveillance

II. Inequitable access

III. Data protection risks

IV. New 'metaverse' risks

V. Useability challenges

VI. Inadequate user support

VII. An untrustworthy ecosystem

Note: This is not meant to be an exhaustive risk assessment — the complete risk universe includes many "real-world risks" (which are the same as would be created by any national ID system), a wide range of security risks (which are true for all aspects of our digital lives), and various identity theft-related risks (which are largely extant today). Instead, the scope here is the *risks to individuals that might arise or increase from a poorly executed online identity assurance framework*.

I. Covert Government Surveillance

"Identification...increases the government's power over individuals."[56] This power has been understood and vigorously debated ever since the first national ID cards were developed in nineteenth century France. As we potentially enable government identification of citizens in their digital lives, we raise new

[56] Understanding Privacy, Daniel Solove, 2008 p 125

risks, especially online surveillance, social control, and political repression. Here are some of those risks and potential safeguards to reduce the risk of mass surveillance by governments.

I.1 Covert Government Surveillance

Risk: Governments may surreptitiously surveil their citizens' (or other users') online activity.

Description of Need: Users require technical protection against government surveillance of their online activities.

Potential Mitigation(s):

- *Support of privacy-protective wallets:* As noted in the previous section, wallets are a cornerstone of empowering users to protect user privacy by shielding online authentication activity from the issuing government (i.e., bypassing any centralized gatekeeper, and not revealing any history of authentication to them). Governments can develop and issue their credentials in support of wallet standards; other countries and NGOs can potentially provide wallets to people in countries with repressive political systems to protect them from their national government.

- *Avoid centralized authentication registries:* Centralized authentication registers (stored by either governments, authorized trust service providers, or OSPs) can be hacked or subpoenaed to reveal the online activity of identified individuals. These should be avoided by design. If governments do insist on keeping this registry, it should be encrypted to reduce the risks from a breach and have only a short retention period. Independent parties should ideally monitor governments, trust service providers, and OSPs, and publicly disclose whether their registers are creating risk.

- *User choice:* Another safeguard against a repressive (or inept) government is to give people choice of who issues, manages, and validates their credentials. Citizenship would not change, but citizens of countries with repressive political systems might authenticate themselves to a trusted intermediary, who then would provide privacy-protective credentials so the user could act freely online without fear of surveillance by their government.

I.2 Legalized Government Surveillance

Risk: National governments may give themselves legal permission to surveil their citizens' (or other users') online activity, and from that surveillance, systematically repress certain people or groups and/or deny them services.

Description of Need: Users (and indirectly, OSPs) require protection from legalized government surveillance of online activities.

Potential Mitigation(s):

- *Define legitimate activity:* Define lawful government interests and activities and incorporate them into an industry code of conduct (see Section 2.4).

- *Embrace international frameworks:* International agreements tend to reflect more moderate government positions, so countries operating within international frameworks (set out by US Cloud Act of 2018[57], for example) might be granted a certain level of trust.

- *Establish a uniform standard of "trustworthiness":* The assessment of relative "trustworthiness" of a government might be made by an independent oversight body representing nations with strong norms around rule of law, and those which adhere to the Universal Declaration of Human Rights, for example. This trustworthiness should determine the degree to which OSPs share information with a national government.

I.3 Erosion of Anonymity

Risk: Despite anti-surveillance measures and legal prohibitions, governments, corporations, and bad actors may still identify users by combining device and network data.

Description of Need: The most strategic safeguard against surveillance would be to legally define and technically implement a (limited) right to anonymity. Presumably, this would establish end-to-end anonymity requirements across devices, networks, applications, and platforms. Anonymous users would rightly have low trust accorded them by OSPs, and so might receive only a limited set of services, commerce might be curtailed, and the ability to post on OSPs services restricted. However, users would be able to access some information and communicate via some channels with assurance that they are not being watched and identified. Individuals might reasonably need to confirm that they are a "real person" (not a bot) and above a certain age — the sort of credentials that can be delivered anonymously via wallets.

Of course, this is a somewhat idealistic suggestion. It would take years of deep technical standards-setting and political wrangling to implement but is important enough to warrant long-term consideration.

Potential Mitigation(s):

- Legal protections to consider include:
    - Requirements for all relevant technology providers across the open systems interconnection (OSI) stack to support an "anonymous mode" of Internet access.
    - Restriction of any tracking, logging, identification, or profiling of individuals in anonymous mode.
    - Restrict any non-consensual identification of users.
- Technical protections to consider include:
    - Wallet support for anonymous confirmation that user is human and above a certain age (all part of selective disclosure capabilities planned for major standards).
    - Cross-stack definition of anonymity rights and requirements such as Dynamic Host Configuration Protocol (DHCP) reassignment, cookie purges, anonymous application instantiation, and encrypted communications.

[57] *The Cloud Act: Strategic Technologies Blog* (no date) *CSIS*. Available at: https://www.csis.org/blogs/strategic-technologies-blog/cloud-act (Accessed: April 19, 2023).

- Technical blockers to prevent any tracking, logging, or profiling of individuals in anonymous mode.
- Operational requirements for TOR improvements (better access, speeds, etc.) to allow anonymous browsing and communication via the TOR network.
- Formalized concept of "identity escrow" where identity information can be shared with trusted service providers, but not revealed to them unless certain criteria are met (execution of smart contract, legal issue raised, terms of service violations, etc.).

I.4 Ambiguous Government Access Needs

Risk: Governments have legitimate reasons to request identity information from OSPs. Policing, investigations, and the prosecution of criminals and terrorists legitimately require sharing of identity credentials. However, police may stretch those legitimate needs into activities that cross the line into surveillance.

Description of Need: OSPs need clear guidelines about what constitutes "legitimate" information requests.

Potential Mitigation(s):

- *Define legitimate government access needs:* Legitimate government interests and activities should be defined and agreed upon as part of an Industry Code of Conduct, along with protocols for requesting that data (if not already supported by the 2018 Cloud Act). Escalation path to other governments should be established if not in place.

II. Inequitable Access

Identification is also a question of equity. Many of the world's poorest individuals have no government-issued identification,[58] which restricts their access to financial services, government help, education and information services. As USAID states, "There may be no single factor that affects a person's ability to share in the gains of global development — to receive services and be represented — as much as having an official identity."[59]

II.1 Inequitable Access

Risk: The most vulnerable people in any country may not be issued government identification and may thus be de facto barred from educational, commercial, and financial opportunities available online.

Description of Need: All global citizens need adequate credentials to access online services and the digital economy, even in a limited manner. Credentials and authentication technology need to be provided at zero (or almost zero) cost.

[58] See the McKinsey study, Digital ID: A key to inclusive growth and the World Bank Study Inclusive and Trusted Digital ID Can Unlock Opportunities for the World's Most Vulnerable (worldbank.org)
[59] Digital Identity | Digital Development | U.S. Agency for International Development (usaid.gov)

Potential Mitigation(s):

A full analysis of potential solutions is out of scope, but a consortium of stakeholders could support the identification of unidentified individuals in several ways:

- *Accept alternative credentials:* Numerous non-profits and NGOs are experimenting with ways to provide trustworthy credentials to the unidentified (UN ID2020[60], ICRC[61], USAID[62], Kiva[63], etc.). Global OSPs should work with these entities to establish a set of acceptable alternative credentials, some of which may have lower associated confidence but should warrant provision of basic services nonetheless.

- *Standardize levels of assurance:* For alternative credentials to be effective, OSPs need a fair and transparent system of scoring credentials for trustworthiness. Access to certain services might reasonably be denied if the confidence in the IDs is too low but should allow this population access to basic and low-risk services.

- *Standard path to increased trust:* The framework should provide a clear, publicly available path by which these people can strengthen the trustworthiness of their credentials. For example:

    - *Third party vouching* – Personal connections who are already verified can vouch for an unidentified person, increasing the trustworthiness of their credential. Aadhaar is using this model, and governments or authorized trust platforms could potentially issue digital credentials to these otherwise unidentified people, which can be used elsewhere (online or even offline).

    - *NGO credentialling and confirmation* – Certified NGOs, who have established their own credentials on the platform, could aid this process by issuing and/or confirming identities, potentially vouching for people as above. This might be combined with biometric credentialling, to serve refugees and other populations in motion, but would ideally be defined by global standards that could be supported with simple tools like smartphones that can strengthen credentials at low cost.

- *Public-private partnership:* The framework should encourage public-private partnerships, which might materialize in several ways:

    - Developing universal standards for "alternative" credentials and/or cooperatively developing trust algorithms.

    - A potential trusted service organization might securely host cloud-based credentials for trusted NGOs, and/or authenticate users online.

    - Online platforms could potentially share user information (with informed consent) to trusted governments and NGOs to help them develop and maintain citizen registries, contact citizens, conduct censuses, and so forth.

---

[60] *ID2020 | Digital Identity Alliance* Available at: https://id2020.org/ (Accessed: April 19, 2023).
[61] *Digitharium* (2022) *International Committee of the Red Cross*. Available at: https://www.icrc.org/en/digitharium (Accessed: April 19, 2023).
[62] *Identity in a Digital age: Infrastructure for Inclusive Development: Document* (2022) *U.S. Agency for International Development*. Available at: https://www.usaid.gov/digital-development/digital-id/report (Accessed: April 19, 2023).
[63] *A retrospective on Kiva Protocol Kiva Protocol*. Available at: https://www.kiva.org/protocol/ (Accessed: April 19, 2023).

- *Ensure universal accessibility:* The ID authentication framework needs to be usable by all people, regardless of age, language, disability status, and literacy. Simple graphical design and video support will be important.

III. Data Protection Risks

These risks concern the misuse of identity information by an OSP. Many of these risks are already addressed by privacy laws in some countries, but since other countries have no privacy legislation or inadequate safeguards within their privacy laws, we consider them significant risks to those users.

III.1 Excessive Identity Data Collection

Risk: OSPs may collect more user identification information than needed.

Description of Need: Users should only be required to share the minimal amount of identifying information necessary.

Potential Mitigation(s):

- *Data minimization policy:* OSPs should review policies and technologies to minimize the credential information they collect; the guiding principle in minimization should be to protect user anonymity to the maximum degree possible.
- *Standardized assurance levels:* OSPs should agree on standard levels of assurance required to access services, and not collect more than required.
- *Selective disclosure:* Governments should provide credentials and support wallets and authentication standards which support selective disclosure and other privacy-protective authentication methods (for example, either ISO 18013-5/7[64] or W3C VC[65]).
- *"Identity escrow":* A more futuristic solution might involve a trusted intermediary who would collect and hold verified credentials and only share with an OSP if there was a violation of terms of service.

III.2 Inadequate Data Subject Rights

Risk: Users may have no way to access, delete, amend, or otherwise control their identifying data.

Description of Need: Users should have full "data subject rights" to their identity credentials and authentication records. These should be easy to exercise and available at no cost to the user.

Potential Mitigation(s):

- *Universal data subject rights:* For identity credentials, OSPs should extend full (GDPR) data subject rights to all users, regardless of jurisdiction. The most relevant rights are:

[64] *ISO/IEC AWI TS 18013-7* (2021) *ISO*. Available at: https://www.iso.org/standard/82772.html (Accessed: April 19, 2023).
[65] *Verifiable credentials data model V1.1* (no date) *W3C*. Available at: https://www.w3.org/TR/vc-data-model/ (Accessed: April 19, 2023).

- *Right to access* – Ability for the user to see what people and/or entities have ID-authenticated them, and via which credentials.
- *Right to rectification* – A user should be able to update their name, contact info, address, gender, and other personal attributes (ideally via their wallets, one time, driving changes across all OSPs with whom they have shared any identifying data)
- *Right to erasure* – A user should be empowered to require the verifying entities to delete any identifying information provided (if it does not breach the terms of service or a contract they wish to uphold).
- *Right to restrict processing* – A user should have the ability to restrict how identifying information is used (though that may come with restriction on the services provided).

III.3 Unintended Use of Credentials

Risk: OSPs may use identity credentials for purposes the user does not expect.

Description of Need: Users should be guaranteed their credentials are not used for unexpected purposes.

Potential Mitigation(s):

- *Purpose limitation restrictions:* Regardless of jurisdiction, platforms should treat credentials as personal data, and strictly limit use to the purpose for which it was collected.
- *Consent requirements:* Consent should be mandatory for alternative usage.
- *Non-transferability:* The verifying party should be restricted from transferring identity credentials to third parties without express consent from the user.

III.4 User Confusion

Risk: Users may not understand how their identity data is being used.

Description of Need: Users should have complete details about the purposes, means, risks, and their rights regarding identity assurance made available to them at each step in the process. Estonia has a system for its citizens, called Personal Data Usage Monitor.

Potential Mitigation(s):

- *User-centric transparency guidelines:* OSPs should adopt a code of conduct specifying transparency requirements regarding all aspects of identification. The World Bank[66] has developed privacy-by-design guidelines,[67] which might be a helpful starting point.
- *Expanded user notice:* OSPs should enhance their user notice to include:

[66] *Home: Identification for development Home | Identification for Development*. Available at: https://id4d.worldbank.org/ (Accessed: April 19, 2023).
[67] PrivacyByDesign_112918web.pdf (worldbank.org)

- The entities with whom they are sharing their credentials.
- The purposes of the authentication.
- Whether the authentication is persistent or transient, and credential retention policies.
- The risks of authentication (e.g., whether their identities might be shared under police subpoena, etc.).
- The retention rules and duration for their identity information.
- User options to see the history of their ID authentication with that entity.
- The options for how users may share credentials or otherwise authenticate.
- Ways for users to exercise their data subject rights.
- The policies by which the information will be processed, especially including any situations when credentials would be passed to third parties.

- *Accessibility guidelines:* Details should be easily accessible, and simple enough to support users of all reading levels, ages, languages, and technical competence.
- *Availability:* Relevant information and links to more detail should be easily available in relevant interfaces of the authentication process.

III.5 Biased Automated Decision Making

Risk: Users may be subject to unseen biases from trust algorithms, namely the algorithms which confirm or determine levels of assurance based on the caliber credentials.

Description of Need: Anticipating that users from different countries and different socioeconomic backgrounds will have different types of credentials available, OSPs will need a mechanism to score the trustworthiness of those credentials and their confidence that an individual has been accurately identified. This mechanism needs to be fair, public, and uniformly adopted as a method to score trustworthiness of credentials.

Potential Mitigation(s):

- *Algorithmic transparency:* Any algorithm which significantly impact users (especially ones that determines the confidence a platform will have in the identity of an individual, based on credential type, etc.) should be made transparent to users, civil society groups, regulators, and other oversight bodies.
- *Standardized trust algorithms:* The industry should consider developing a standard trust algorithm to ensure uniformity.

IV. New "Metaverse" Risks

The potential risks of the metaverse are more theoretical than actual, as the concept of a "metaverse" is still evolving, and therefore many other risks might emerge. But broadly, it is not hard to imagine

scenarios whereby it becomes a disorienting and manipulative experience, with significant potential harms. Having a clear metaverse identification protocol can alleviate some of those risks.

IV.1 Bot Misrepresentation

Risk: It is likely that bots will soon be indistinguishable from humans via chat, video, or audio channels. They might then misrepresent themselves as personal assistants of individuals, employees of companies, or other individuals, causing confusion or manipulation.

Description of Need: All parties should be entitled to know if they are engaging with a bot, and on whose behalf the bot is operating.

Potential Mitigation(s):

- *Bot identification policies:* OSPs and other bot enablers should enact policies, implement safeguards, and offer tooling to ensure that all bots are proactively identified, including the fact that they are a bot and the entity on whose behalf the bot is operating (i.e., "the bot sponsor"). OSPs should make bot identification policies explicit as part of their Terms of Service.
- *Bot sponsor accountability:* Bot sponsors should be held accountable for bot actions.
- *Bot identity transferability:* If bots are eventually able to move freely between metaverses (to serve one entity in multiple platforms), standards should ensure that credentials and identification requirements move with them.

IV.2 Asymmetric Identification

Risk: In a decentralized, multi-stakeholder environment, some parties may obtain identifying information without identifying themselves, or share less information than they are getting about a particular individual.

Description of Need: Whenever a user shares a validated credential, by default the verifying party (whether a business, government, or person) should also share a similar credential of at least the same level of detail ("symmetry"), particularly if the user might have any doubt about their ID.

Potential Mitigation(s):

- *Symmetric identification policy:* OSPs should adopt policies requiring identification symmetry whenever possible on their services.
- *Standard ID detail scale:* Supporting this, OSPs would need to agree on a scale of "credential detail" to define the concept of symmetry in practice.

IV.3 Failed Recall

Risk: In multi-stakeholder environments, a user might easily authenticate their ID with a person or business, and subsequently forget they have done so. This could be common where third parties (e.g., consumer brands) exist in multiple metaverses and under different brand names. Users might not

realize that those entities have visibility into their identity from another metaverse. The business could leverage automated recognition, so there would be an asymmetry in the relationship.

Description of Need: When a user meets a previously verified entity (even if from another metaverse or under a different brand):

- The user should be reminded that the other party knows their identity.
- The other party's identity should be shared.

Potential Mitigation(s):

- *Persistent acknowledgement:* OSPs should enact policies, implement safeguards, and offer tooling to support persistent acknowledgement of identification.

IV.4 Entity Misrepresentation

Risk: In addition to users, millions of online businesses, religions, government agencies, and other entities will likely operate in "metaverse ecosystems." Without a standard authentication protocol, users may be defrauded or manipulated by unscrupulous entities.

Description of Need: Businesses and other entities will need to authenticate themselves, their employees, and eventually their bots via the same standards and protocols that OSPs will use elsewhere.

Potential Mitigation(s):

- *Business authentication protocols:* Leading OSPs should accommodate technical and legal standards that require businesses to appropriately authenticate themselves, and their agents (human and digital), to users.
- *Advertising accountability:* Metaverse advertisements should be explicitly associated with an authenticated business, and that association should be easily visible to all users.

IV.5 Confusing "Cross-Credentialling"

Context: Metaverses will likely be complex, multi-stakeholder environments. Numerous businesses, artists, and other entities will operate independently, and those entities themselves might each have numerous autonomous agents (people, bots, assistants, etc.). Users also will be multi-faceted, potentially having multiple accounts and personas inside a given metaverse.

Risk: These will create complex credentialling questions, which may leave users unclear about where and when they have been identified. For example, once a user has authenticated with a platform, to what degree should the platform authenticate them to other services and businesses supported within a platform? Should the user be given control and responsibility to manage authentication each time? How should OSPs enable peer-to-peer authentication (ideally without the OSPs in the middle of the transaction)? When a user moves from Metaverse A to Metaverse B, must they reauthenticate with each of the same entities? What if a user has multiple accounts with the same OSP? Should the credential apply to all accounts?

Description of Need: OSPs should establish clear guidelines on appropriate "ross-credentialling" before development begins.

Potential Mitigation(s):

- *Rules for cross-credentialling:* Platforms should establish guidelines governing the forwarding of credentials between business entities and subsidiaries, across users' profiles and accounts, and between metaverses.
- *Cross-credentialling transparency:* As this may be confusing to users, and there may be significant privacy harms, special care should be given to ensuring users understand how their identity information is being shared between their own accounts and between authenticated contacts.

V. Useability Challenges

Users may make mistakes, over-share identity information, or abandon a service altogether if the identity assurance process is not straightforward and easy to use.

V.1 Non-Standard ID Assurance Protocols

Risk: Users are more likely to make mistakes in their authentication (authenticate with entities erroneously, over-authenticate, etc.) if they are required to use different tools, different interface designs, different processes, or different protocols each time they authenticate. (Such friction will also decrease the likelihood that a user will join or use an OSP service.)

Description of Need: Principles of fairness suggest that processing must be done in ways that people would reasonably expect.[68] In order to consistently meet user expectations, OSPs should offer a standardized and easy-to-use authentication process.

Potential Mitigation(s):

- *Standardized user experience:* OSPs should jointly develop industry-standard processes with standard user notices, interface designs, terminology support, and graphical explanations. Digital wallet standards should interoperate with these standards as they are adopted.
- *Interoperability:* Ensure that user credentials, wallets, and authentication techniques are accepted by all platforms, with the same terms. Development of multiple regional standards might be practical. Standardized assurance protocols should apply to authentication via multiple types of devices.

V.2 Unachievable Credential Requirements

Risk: User ID authentication may be denied because OSPs do not accept their credentials

[68] The European Data Protection Board (EDPB) has defined a useful set of fairness principles (*EDPB Fairness Design Elements - European Data Protection Board (n 132) 65)* and one of them, 'Expectation', states that "Processing should correspond with data subjects' expectations"

Description of Need: Users need clarity on what credentials are required for certain levels of assurance, and the options they have for alternate credential verification. This protocol should be standard across OSPs. Different types of credentials would reflect different trust levels (per NIST SP 800-63-3[69] or other standards).

Potential Mitigation(s):

- *Credential protocol*. OSPs should develop a standard credential protocol, establishing acceptable credentials for each country (whether issued by the national government or other approved entities). Within a given level of assurance, OSPs should maximize the range of accepted credentials to maximize participation.

VI. Inadequate User Support

VI.1 Inadequate Front-Line Support

Risk: In the event of an event that requires user support (especially a complex one such as identity theft, a user's death, or device theft), a user may get stuck between the support systems of various entities — credential issuers, device makers and their distributors, applications, intermediaries, OSPs, outsource providers, and so forth.

Description of Need: Users need clarity on which entities will offer different types of support, and ideally the support should be coordinated end-to-end across different entities if needed.

Potential Mitigation(s):

- *Defined support model:* Governments, OSPs, and any potential intermediaries need to define a clear support model, including channels of accountability, easily accessible contact points, operational interlocks in the event of complex problems (e.g., identity theft), and service-level agreements (SLAs) with appropriate reporting and enforcement.
- *User notice of support:* The support model should be communicated consistently to users.
- *Multi-channel support:* Support should ideally be provided via multiple channels (phone/video, email, chat, in-person, etc.) and be consistent across countries, OSPs, and technology providers if possible.
- *SLA oversight:* Adherence to SLAs should be overseen by an oversight body.

VI.2 Government-Issued Credentials Not "Internet-Ready"

Risk: Government-issued credentials may be non-digital, may be incompatible with digital wallets, or may lack other privacy-protective features.

Description of Need: Users depend on governments to issue Internet-ready credentials. If they are not already digital and privacy-protected, the government should ensure interoperability with digital wallets

[69] *NIST SP 800-63 Digital Identity guidelines* (no date) *NIST SP 800-63*. Available at: https://pages.nist.gov/800-63-3/ (Accessed: April 19, 2023).

as those emerge, but there will be lagging governments as well as individuals who will have unprotected credentials for the next few decades.

Potential Mitigation(s):

- *Suggested standards:* As ID assurance standards mature, and if service providers embrace a code of conduct, standards to define 'Internet-ready credentials' will help governments develop a roadmap, and should include workarounds to privacy-protect existing analog credentials (such as the Aadhaar model, which issues a unique virtual ID on top of the physical ID).
- *Intermediary services:* The framework could certify or offer trusted intermediary services to "mask" analog IDs, offer a "cloud wallet" (useful for populations without regular access to phones), or other workarounds. The intermediary might need to authenticate a user and then issue derived credentials which OSPs could then accept.

VII. Untrustworthy Ecosystem

Maintaining mutual trust between users, governments and OSPs is arguably the most important factor in the long-term effectiveness of online identity authentication. Governments and OSPs need to prove their trustworthiness to each other, and to the individuals they serve, for the system to function. If any party abuses identity information, the entire framework will be threatened.

7.1 Weak or Dispersed Accountability

Risk: Users may be harmed by OSPs or governments who do not comply with responsible standards; users may not embrace online services if they do not trust the identity authentication protocols provided.

Description of Need: Users, governments, and OSPs need to trust the security, utility, and fairness of the authentication system. Establishing trust is a primary responsibility of governments and OSPs.

Potential Mitigation(s):

- *Oversight body:* OSPs and leading regulators should create an independent oversight body to ensure a trustworthy approach to identity assurance. This body might monitor governments, trust service providers, and OSPs for compliance with their agreed responsibilities and policies and then publicly disclose findings. The body should also agree on a trust methodology, likely including:
    - Establishing necessary governance structures and rules for participation from all stakeholders.
    - Agreeing on a regular, public reporting cadence to all stakeholders and inviting external stakeholders to review its progress.
    - Establishing independent audit protocols, and ensuring members comply with audit requests.

- Creating a certification mechanism to acknowledge and protect OSPs who adhere to best practices, trusted intermediaries, technology providers who meet technical standards, and governments who support fair authentication.
- Developing a mechanism to constrict toxic platforms, for example by coordinating an industry-wide boycott (devices, applications, and service providers) in response to harmful behavior.

- *Transparency:* Countries which are not the principal overseers should have full visibility into the guidelines and operational measures used to process these credentials and be regularly provided with country-specific operational reports. In case of disputes, the oversight body should offer governments an escalation path to raise concerns.
- *Advocacy for privacy:* The oversight body should advocate to governments to ensure they are issuing privacy-forward Internet-ready credentials as soon as possible. These privacy-protective features might include:
  - Credentials which support selective disclosure and other privacy-protective assurance features.
  - Credentials are accepted by mDL (ISO18013-7[70]) wallets or W3C[71] verifiable credential wallets.
  - Hashed identifiers to enable a verifiable virtual ID number.
- *Solvency:* For continued oversight, this body needs to remain solvent long-term, so should ascertain sources of its operating budget in its creation

7.2 Technical Lock-In

Risk: Users and governments will be harmed if the authentication market is captured by a select group of technology vendors or verifying parties, which could result in technical lock-in, lack of innovation, poorer security, and higher prices.

Description of Need: Governments should advocate for open standards, as opposed to proprietary standards, to enable safe innovation and continual improvement while avoiding vendor lock-in.

Potential Mitigation(s):

- *Open standards:* The industry should embrace open standard for wallets, credentials, trust assessment algorithms, authentication tools and frameworks, and all supporting technologies.

[70] *ISO/IEC AWI TS 18013-7* (2021) *ISO*. Available at: https://www.iso.org/standard/82772.html (Accessed: April 19, 2023).
[71] *Verifiable credentials data model V1.1* (no date) *W3C*. Available at: https://www.w3.org/TR/vc-data-model/ (Accessed: April 19, 2023).

This table summarizes these risks and mitigations in brief:

Societal Risks from Identity Assurance			
Category	**Num**	**Risk**	**Potential Mitigations**
Government Surveillance	1.1	Covert government surveillance	Architectural safeguards Privacy-protective wallets Avoid centralized authentication registries User choice
	1.2	Legalized government surveillance	Define legitimate activity Embrace international frameworks Establish a uniform standard of “trustworthiness”
	1.3	Erosion of anonymity	Legal anonymity protections Technical anonymity protections
	1.4	Ambiguous government access needs	Define legitimate government access needs
Inequitable Access	2.1	Inequitable access	Accept alternative credentials Standard levels of assurance Standard path to increased trust Public-private partnership Ensure universal accessibility
Data Protection Risks	3.1	Excessive identity data collection	Data minimization policy Selective disclosure Identity escrow
	3.2	Inadequate data subject rights	Universal data subject rights
	3.3	Unintended use of credentials	Purpose limitation restrictions Consent requirements non-transferability
	3.4	User confusion	User-centric transparency guidelines Expanded user notice Accessibility guidelines Availability
	3.5	Biased automated decision making	Algorithmic transparency Standardized trust algorithms
New 'Metaverse' Risks	4.1	Bot misrepresentation	Bot identification policies Bot sponsor accountability Bot identity transferability
	4.2	Asymmetric identification	Symmetric identification policy Standard ID detail scale
	4.3	Failed recall	Persistent acknowledgement
	4.4	Entity misrepresentation	Business authentication protocols Advertising accountability
	4.5	Confusing “cross-credentialling"	Rules for cross-credentialling Cross-credentialling transparency

Inconsistent Standards	5.1	Non-standard ID assurance protocols	Standardized user experience Technical interoperability
	5.2	Unachievable credential requirements	Credential protocol
Inadequate User Support	6.1	Inadequate front-line support	Defined support model User notice of support Multi-channel support SLA oversight
	6.2	Government-Issued credentials not “Internet-ready"	Suggested standards Intermediary “masking”, “cloud wallet” or other services
Untrustworthy Ecosystem	7.1	Weak or dispersed accountability	Oversight body Advocacy for privacy Solvency
	7.2	Technical lock-in	Open standards

V. The Path Forward

Transitioning to Privacy-Protective Identity Assurance

In the previous section, we identified 7 categories of risk and 51 potential safeguards to maximize the societal benefits and minimize potential harms of identity assurance. In this section, we discuss possible next steps to begin responsible design of a solution.

Identity assurance is inherently a challenging topic:

- A managed solution will require coordination between governments, technology and standards organizations, and online service providers.
- It impacts all countries, so a solution will require alignment between at least a few governments; there is currently no regulatory body managing problems like this.
- Technology companies and online platforms are competitive in nature and distrustful of regulation, which makes cooperation difficult to achieve.

Given these challenges, some would say it would be better to wait a few years and see how the market unfolds before trying to align such a complex set of interests.

What If We Do Nothing?

What if we did that? What if governments fail to take action on this topic and if companies make no attempt to coordinate? 'Doing nothing' is always an option we should consider.

Here is the author's prognosis in that scenario:

In a more positive scenario:

- Some standards-based solution will likely emerge as dominant and gain widespread acceptance in some countries.
- Some governments will support their citizens with internet-ready credentials and restrict police surveillance
- Some service providers will act responsibly, and leading players may develop workable standards for interoperability.[72]

  But it is unlikely that these positive outcomes will extend everywhere, so we will create a patchwork or solutions leading to global inequity. The lack of clear rules of conduct will open the door to abuse by some companies. Some governments will pass their own laws requiring users to authenticate at much higher levels of assurance than others, thus violating their privacy. Etc.

In a more pessimistic scenario:

[72] The Metaverse Standards Forum, for example, includes identity as one of its 5 working groups, though scope and outcome are still to be determined

- Private companies will view identity assurance as a business opportunity; they aggressively try to capture market share with a proprietary solution (possibly using AI to validate credentials and create economies of scale).
- This may solve the short-term problem of bot identification but may also lead to:
    - Technology lock in
    - Market power abuses
    - Global inequity
    - Loss of control by governments over their citizens identity protection
    - Any of the other risks outlined above may also be realized

One Possible Path Forward

Given the possibility of so many undesirable outcomes, it seems reasonable that governments should try to take action. If they so choose, a public-private partnership seems like the most promising model. A purely regulatory approach is challenged by the lack of a single global regulator and the concern that technology is evolving so fast that regulations risk becoming obsolete soon after passage. A purely technological approach (i.e., the establishment of standards and free market development of compliant solutions) would leave users vulnerable to government surveillance, corporate data protection abuses, and unclear support models.

The ecosystem thus needs to come together to begin framing a solution. This initial focus should be on establishing a joint understanding of values, risks, and potential mitigations. This would likely start with a handful of regulators from the most technically advanced countries, leading technology players, and a few of the largest tech companies. Over the next few years, one could envision the establishment of an "Online Identity Assurance Framework" via a public-private partnership. This Framework might logically be comprised of five functional areas:

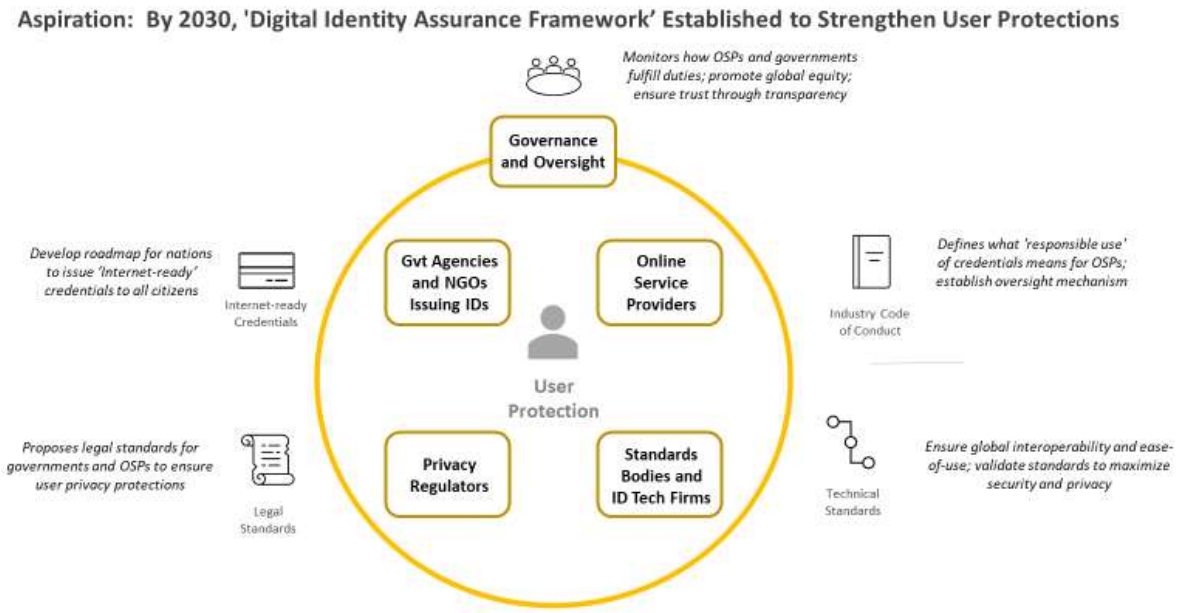
1. *Governance and Oversight Body:* A joint public-private oversight structure to ensure high levels of transparency, accountability and trust.

2. *Online Service Providers:* A combination of social media, financial services, governmental and other digital service providers that would develop and operate under an Industry Code of Conduct that would clarify the principles and guidelines by which to manage user credentials.

3. *Standards Bodies and ID Technology Firms:* Organizations that would develop interoperable Technical Standards and products to ensure security, useability and privacy protections for wallets, credentials, and other pieces of a credential verification system.

4. *Privacy Regulators:* Regulators that would establish legal standards to ensure government and corporate surveillance risk is curtailed, and over time, to codify restrictions deriving from the OSP Code of Conduct or other regulatory innovations.

5. *Government Agencies and NGOs Issuing ID Credentials:* Credential issuers that would develop country-specific plans to ensure Internet-ready credentials are available to all individuals; in countries with inadequate government support, NGOs could issue authorized credentials in lieu of government provision.



Here is a mapping of some of the key risks which the functions would need to address:

1. The *Governance and Oversight* function would be tasked with ensuring the overall trustworthiness of the Framework, and its security, utility and fairness. It would likely take the form of a committee or board, and would need to:

    - Establish necessary governance structures and rules for participation from stakeholders.
    - Develop independent audit protocols and regular public reporting.
    - Create mechanisms to recognize good behavior and penalize harmful behavior.
    - Ensure clear accountability for user support across governments, intermediaries, and OSPs.
    - Support efforts to minimize global inequity, including expanding allowable credentials.
    - Ascertain sources of the Framework's operating budget to ensure its continued functionality.

2. The *Industry Code of Conduct* should develop guidelines to ensure OSPs use credential information responsibly. In the short term, the focus should be on addressing any current data protection risks. Note that most of these protections are already required for GDPR compliance, so many OSPs will have implemented them (at least for European users). In the short term, it should be relatively simple to extend these requirements to rest-of-world users and enshrine them as a first-version code of conduct. Requirements might include:

    - Minimize credential data collected.
    - Maximize user control with regard to authentication procedures.

- Enforce purpose limitation protections on identification data collected.
- Offer full Data Subject Rights to all users (for credential data).
- Standardize set of acceptable credentials across OSPs, ensuring they are reasonable, and support varying levels of assurance as required.
- Offer (standardized) authentication protocols including standard notices, help, and escalations.
- Establish transparency guidelines for authentication and the processing of identification data.
- Refine algorithmic transparency practices for algorithms evaluating identity credentials.

The remaining topics are mostly metaverse-related so may take more time to refine and agree upon. These potentially include guidelines concerning identity authentication in the metaverse such as:

- *Bi-directionality and Symmetry*, requiring third parties to identify themselves at the same level of detail as the user when they authenticate themself.
- *Persistent Acknowledgement Protocol*, ensuring that users are informed when they are in contact with an entity who has already authenticated their identity.
- *Bot Identification Protocols*, establishing rules for how bots identify themselves, including the person, business, or other entity they are serving.
- *Metaverse Business Authentication*, setting rules for how businesses are authenticated and how they identify themselves to users.
- *Rules for Cross-Credentialling*, ensuring transparency and user control of how authentication is transferred between user accounts and profiles and with complex entities spanning multiple metaverses and/or with multiple supported brands.

3. The *Technical Standards* area would improve ease-of-use, security, interoperability, and privacy protection in general. The function would work to:
    - Ensure increasing interoperability between different trust standards over time.
    - Push technology firms towards greater ease of use and authentication across multiple devices.
    - Design technical counter-surveillance protections.
    - Advocate for development of broader anonymity protections.
    - Ensure end-to-end security for users and their identifying data.

4. The *Legal Standards* team might establish legal standards (which would need to be adopted by national regulators) to:

- Prohibit non-consensual identification.
- Legally protect users against government or commercial surveillance.
- Strengthen anonymity protections through proposed statute.

5. *Credential-Issuing Bodies* would develop a roadmap to:
    - Ensure governments are designing their national credentials to be as 'Internet-ready' as possible.
    - Minimize global inequities of ID assurance through NGO participation and acceptance of alternative credentials.
    - Help individuals understand, embrace, and use new technology made available.
    - Ensure individuals understand the risks of ID assurance, as well as their rights.

Who Will Take the First Step?

To many, coordinating such a diverse set of global stakeholders to manage rapidly developing technologies may sound naïve. Indeed, this is not a simple effort, but any journey begins with a few small steps and minor milestones. Many stakeholders are in fact already moving ahead and aligning on important elements of the solution:

- Several standards bodies (ISO 18013-7, W3C VC, eIDAS 2, NIST, etc.) are already well organized and developing important standards for ID assurance. Each of these standards bodies have embraced user privacy protections as cornerstones of their technical strategy.
- Technology companies such as Apple, Google, Microsoft, and many smaller firms are building the products and solutions needed for identity assurance.
- Large service providers also recognize the importance of ID assurance and are self-organizing via the Metaverse Standards Forum[73]. Progress in this forum is nascent however, and scope is still evolving.
- The OECD recently issued guidance[74] recommending that member governments "take a strategic approach to digital identity and define roles and responsibilities across the digital identity ecosystem."
- The EU Commission has built alignment with 27 EU countries to establish the European Digital Identity Framework as part of its Digital ID Act. This includes requirements for national governments to produce compliant credentials, technical standards for the EU Digital ID wallet, and a series of regulations ensuring that the new credentials can work seamlessly across borders and meet both online and real-world needs.

[73] https://metaverse-standards.org/
[74] Draft OECD Recommendation on the Governance of Digital Identity (Public Consultation Version) (oecd-opsi.org)

In short, many of the primary stakeholders are already wrestling with critical parts of this issue. What is missing is a coordination function to bring these moving parts together. The United States government (aside from NIST) is conspicuously quiet on this issue, lacking both a national ID strategy and a regulatory approach, not to mention any broader privacy protection regulation. But if the US government were prepared to engage, a simple way to begin this coordination might be:

- Join appropriate regulators from the EU and India[75] to conduct a forum on online identity assurance and try to align on basic principles (e.g., the value chain in Section III), regarding user privacy as a foundational objective.
- If they can agree on directional principles, these founding bodies convene a somewhat larger forum with service providers, standards bodies, technology providers, and NGOs to discuss these principles. The ambition would be to articulate a simple charter expressing the values and accountability for identity assurance. The charter should lay the foundation for establishing the "Digital Identity Assurance Framework" and lay out requirements for participation.
- Once the charter is defined, industry organizations and firms themselves can consider whether they would support the charter, and if so, make public commitments to it. Failure to adopt the charter might suggest a need for more traditional regulatory oversight on the topic.
- The Governance and Oversight body can be formed (suggested to be a public-private partnership) and the rest of the organizational charter established.

Of course, this is only one of many possible paths that illustrates the phased nature of building alignment. With political will and leadership, a global framework is achievable.

The Benefits of Responsible Identification

Establishing a shared vision for the future is a critical first step. While expanded online identity assurance increases risks, it also potentially brings many benefits, and can serve as a stabilizing foundation in an era of profound change. By 2030, the next incarnation of the Internet could be safer, less fraudulent, and less manipulative. Youth in particular should expect a more protected environment, tailored to their best interests (as defined by national regulators). There are many reasons for optimism about this next era, and an Online Identity Protection Framework will mitigate some of the risks of this transition. Here are the some of the benefits each party can expect from a better model:

Benefits to Individuals

- Improves the safety and integrity of online environments.
- Makes access to online services easier with the elimination of most passwords, and standardized authentication processes.
- Ensures privacy protections via wallets and code of conduct.

[75] Given its size and technical power, China would be an obvious fourth participant. However, it has chosen to make identity verification the foundation of its online surveillance state, so should not be included. Other nations might rightfully protest that they should have a voice in this proceeding. While this is true, the challenges of aligning between the three market leaders are significant enough that initial membership should be constrained. After initial alignment, other countries need to be given a channel to raise their perspectives and concerns.

- Clarifies the support model for identity-related issues.

Benefits to Governments and Regulators

- Improves the general safety and integrity of online environments.
- Establishes oversight and control of their citizens' online identity assurance, without requiring a lengthy regulatory development process
- Accelerates adoption of national digital IDs and wallets, which in turn enable digital government services such as benefits delivery, tax collection, and national service registration;
- Reduces costs of online authentication reduces barriers to entry for new online businesses, easing anti-competitive concerns.

Benefits to OSPs and online businesses

- Gives OSPs a safe, low-cost mechanism with which to build protections against fraud, hate speech, misinformation, and youth exploitation;
- Allows businesses to verify their own identity (and their employee's relationship as representatives of the company) with customers, partners and suppliers, potentially reducing fraud, etc.
- Simplifies and standardizes identity assurance operations with industry-standard guidelines.
- Reduces compliance risk via adherence to an Industry Code of Conduct;
- Eliminates ID as a competitive differentiator, as all competitors have the same obstacles, risks, and cost structures based on the Code of Conduct.
- Streamlines regulatory oversight via adhering to agreed-upon global standards (as opposed to myriad state and national requirements).

A Call to Act

The Internet is entering a period of profound transition, with fast-moving technologies likely to disrupt social and economic systems around the world. The issue of identity assurance is upon us and it is increasingly clear that, with or without a plan, we will soon transition to greater levels of identification. Lacking a plan, we foresee an Internet where identity assurance becomes a global market segment dominated by big tech firms and susceptible to surveillance, abuse and monopoly technical lock-in. At this stage however, as Internet safety is in the interest of all people, companies and governments, it seems reasonable that leading stakeholders can align on basic principles and design relevant safeguards to protect us all from online harms. We should act soon before it is too late.

Appendix

Analysis of EDPB (European Data Protection Board) Design Elements for Fairness

Design Element	EDPB Explanation	Relevance to Identification
Autonomy	Data subjects shall be granted the highest degree of autonomy possible with respect to control	Users should be able to control whether they identify themselves, with whom they share their identity, and what level of details shared
Interaction	Data subjects must be able to communicate and exercise their rights with the controller.	Authentication information included in data subject rights
Expectation	Processing should correspond with data subjects' expectations	Requires standard approach to identification. Users must be aware they are being identified
Non-discrimination	The controller shall not discriminate against data subjects.	Beyond the scope of this review
Non-exploitation	The controller shall not exploit the needs or vulnerabilities of data subjects	Beyond the scope of this review
Consumer choice	The controller should not "lock in" their users. Whenever a service or a good is personalized or proprietary, it may create a lock-in to the service or good. If it is difficult for the data subject to change controllers due to this, which may not be fair.	Beyond the scope of this review
Power balance	Asymmetric power balances shall be avoided or mitigated when possible. Controllers should not transfer the risks of the enterprise to the data subjects	NA
Respect rights and freedoms	The controller must respect the fundamental rights and freedoms of data subjects and implement appropriate measures and safeguards to not violate these rights and freedoms.	Enable DSAR
Ethical	The controller should see the processing's wider impact on individuals' rights and dignity	OSPs must mitigate the risks of enabling a digital police state
Truthful	The controller must act as they declare to do, provide account for what they do and not mislead the data subjects.	NA
Human intervention	The controller must incorporate qualified human intervention that is capable of recovering biases that machines may create in relation to the right to not be subject to ADM in Article 22.	NA
Fair algorithms	Information shall be provided to data subjects about processing of personal data based on algorithms that analyze or make predictions about them, such as work performance, economic situation, health, personal preferences, reliability or behavior, location or movements.	Algorithms which assess the confidence in individual identity should be made transparent to users and oversight bodies.

EDPB Fairness Design Elements - European Data Protection Board (n 132) 65.