On Digital Identity Frameworks and Wallets and their Privacy Harms to Consumers

Sylvain Chatel[1] Christian Knabenhans[2] Mathilde Raynal[2]
Theresa Stadler[3] Carmela Troncoso[4] Shannon Veitch[5]
[1] CISPA Helmholtz Center for Information Security [2] EPFL [3] Swiss Data Science Center [4] MPI-SP [5] ETH Zurich

***Abstract*—Digital identity frameworks (DIFs) are systems that enable users to obtain *authenticated attributes* (electronic identity documents, membership certificates) from issuers (governments, banks, private institutions) and store them in a *digital wallet*. Later, users can present (a portion of) such attributes to relying parties or verifiers (to obtain access to services, to prove eligibility for benefits, etc.). DIFs are being proposed and deployed in multiple jurisdictions worldwide, often backed by regulations. A prominent example is the European Union's proposed eIDAS 2.0 regulation [1], which envisions a broad ecosystem of attribute issuers, relying parties, and use cases. In the US, the development of mobile driver's licenses goes in the same direction. The privacy of these systems (or lack thereof) is an important consideration for consumers, and can hinder their deployment and acceptance [2]. Therefore, much of the discussion around these systems revolves on how DIFs can be built while limiting information flows.**

We argue that focusing the discussion on information flows actually impedes an open discussion of the *harms* stemming from introducing DIFs in our daily digital interactions. The lack of discussion prevents consumers from understanding the implications of using DIFs, prevents researchers from studying solutions that prevent wider harms, and prevents policy makers from producing effective regulations that can protect researchers. To aid in this debate, we identify harms that stem from introducing a digital identity framework, from using it as a tool to present verified attributes to enable digital interactions, and from particular implementation decisions. Some of these harms are inherent to the technology and its intended uses, and thus cannot be removed; and some depend on design decisions and might be mitigated by concrete implementations. We illustrate how these harms can materialize in a preliminary analysis of DIF-based age verification, an application of DIF gaining importance across the globe.

1. Identifying and categorizing harms

To better understand how harms come to be, and how they are related to the information a system must reveal to fulfill its functionality, we introduce a novel methodology based on formal modeling of systems, inspired by the security and cryptography literature. We use an *ideal functionality* to capture parties, interfaces, and interactions present in a DIF; and to capture the high-level security and privacy properties which we extract from the regulatory framework in which it is to be deployed (in this case, eIDAS 2.0). The use of this ideal functionality enables us to separate harms inherent to the system's existence and intended functionality (which thus *cannot be eliminated by any implementation decision*) from harms resulting from concrete implementation decisions and therefore *can be mitigated by better system design*.

Using the leakage the ideal functionality reveals, we derive *harm trees* [3] (a privacy equivalent of attack trees, a methodology widely used in the security community) to systematically explore the potential harms to consumers that may arise from the system and their interrelation. These trees enable us to map how harms we identify stem from components of the system; we then formalize the absence of these risks as desirable privacy properties.

Harms stemming from the digital nature of DIFs. A first class of harms arises from digitalizing the act of showing an ID (and associated attributes). The introduction of a DIF facilitates sharing of attributes authenticated by a trusted issuer. Such attributes are more valuable to data brokers than non-authenticated, self-disclosed data as collected today. This creates an incentive for various parties to request (more) consumers' authenticated data vouched for by a trusted issuer, creating new data flows inexistent in current systems. These new data flows *increase consumers' risk to be profiled and targeted or discriminated against*. Due to the use of digital infrastructure, these attributes are inherently made available to third parties such as internet service providers or ad networks, who might correlate signals from existing tracking mechanisms (IPs, cookies, browser fingerprinting, etc.) with authenticated attributes to track consumers more effectively or more persistently.

Easy access to authenticated attributes might lead to a flare-up of disclosure requests for users accessing offline or online services and a risk of over-identification. This can lead to *increased censorship or self-censorship* and *diminished consumer access to resources*, limiting consumers' ability to collaborate and create content, and *exacerbating structural exclusion*. Exclusion and access restrictions are aggravated by the need for devices to have digital interactions due to devices' ephemerality or requirements on software or hardware not available at large.

Harms stemming from the envisioned functionality. A main functionality of eIDAS-based DIFs, required in many proposed use cases, is to identify users (e.g., by revealing their name, ID, social security number, account number) [4]. This, *by design, enables tracking (and associated harms)*,

and prevents any privacy benefit from selective disclosure mechanisms. The eIDAS 2.0 regulation does not enforce control on what attributes are asked from users, thus it is possible to create pseudo-identifiers from attributes (even if the user's identity is not explicitly requested) which allow for *tracking even without identity information*. This risk exists even if user consent is required, as there is no mechanism to aid users in assessing and rejecting consent when they are asked for attributes that might cause non-obvious harm (e.g., correlated or repeated requests). A second goal of the DIF is to serve as an authenticated source for authorization, enabling *systematic and accurate discrimination and targeting based on attributes*, even if disclosure is made in a privacy-preserving way (e.g., targeting consumers living in an area, or within an age bracket).

Harms from DIF's security and privacy implementation choices. Given the privacy concerns DIFs raise, the eIDAS 2.0 regulation includes a provision for unlinkability (for use cases where identity is not needed) [1]. Using unlinkable credentials with *selective disclosure* as planned in the European Wallet, or in proposals by Google [5] and Microsoft [6], can approach the minimal leakage from the ideal functionality (which limits the harms to the two categories above). However, other jurisdictions (e.g., Switzerland) rely on batch-issued credentials, which enables colluding issuer and verifier (e.g., a government issuer issuing subpoenas to get access logs from service providers) to link attribute presentations, thus exacerbating all of the aforementioned risks and *effectively creating a surveillance system*.

For security reasons, the EU DIF relies on device binding to prevent sharing of wallets across multiple users. This might increase *access and exclusion harms*: for example, multiple users might share a single device (e.g., due to cultural or socio-economic backgrounds [7]). Such harms can also appear as a result of the use of unlinkable credentials (or more advanced PETs) which might impose additional hardware or requirements on devices that users might not have access to; or prefer not to use (e.g., consumers preferring non-Google or - Apple phones), affecting the *free choice of consumers* by creating an unbalanced market.

2. Analysis of DIF-based Age Verification

Age verification is emerging as a desired feature by policy makers and users to protect minors. In this model, consumers would get an authorized age attribute from an issuer such as a state authority, and they would prove this attribute is above or below a pre-defined threshold to gain access to services, e.g., social media, online gambling websites, adult content, or private messaging. In this use case, identity is not mandatory, and it is likely that the same user returns to the same service, so the eIDAS 2.0 requirement for unlinkability applies.

The use of a DIF for online age verification materializes many of the harms above (in the Appendix, we show the ideal functionality, the harm trees, and how the combination leads to understand which harms are independent of the implementation).

First, it might *exacerbate the structural exclusion of certain groups from online resources*: children, adults with low digital literacy, or with no access to official identity documents. Service providers pulling out of territories where the market size does not offset the added costs of complying with age verification regulations might further reduce access to services for consumers from certain geographical regions.

Besides discrimination due to lack of access, the introduction of age verification enables services to *implement discriminatory practices* based on age itself, or through profiling/tracking that is enabled through the DIF. Furthermore, once the technology is in place, services that do *not* need to implement age verification can still opt to discriminate between users by relying on the DIF system.

The introduction of DIF-based age verification might lead to *censorship in multiple ways*. The additional burden of verifying age may result in a chilling effect on legitimate activities (e.g., users not publishing content online or not accessing information such as news), *reducing even more consumer access to resources and undermining freedom of expression*. Concurrently, those deciding which services require age verification can restrict what children can see online and which platforms and services they can use.

We note that all the above forms of access deprivation can be harmful to children, as they impair their self-development (biased educational content) and restrict their freedom of expression (blocked access from private messengers or social media).

Moreover, the use of a DIF for online age verification is set to *further deepen the harms of targeting online advertisements* such as harmful biases, facilitating the spread of disinformation, and exploiting vulnerable populations. Minors can be targeted for advertisements based on verified age information coming from the system. These effects can be increased if the DIF enables the creation of authenticated, traceable pseudoidentifiers for individual consumers.

The introduction of age verification via DIF also has the potential to diminish internet safety, when the associated DIF enables tracking – and even de-anonymization – as it prevents anyone who needs independence from the issuing authority or anonymity online from safely using the internet (such as investigative journalists, activists, or some researchers), and might discourage consumers from expressing unpopular opinions. Moreover, consumers who want anonymity might be pushed to unsafe/unregulated spaces which increases their exposure to risks such as fraud or theft.

3. Next steps

We present a new methodology and a categorization of harms that arise from the deployment of DIFs. A remaining question is whether the categorization and harms found are complete, and whether the privacy harm trees are the best way to comprehensively analyze harms and their origins. It is also important to identify other scenarios in which the methodology can be valuable to assess its generality.

References

[1] European Union, "Regulation (EU) 2024/1183 of the European Parliament and of the Council of 11 April 2024 amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework," 2024. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32024R1183

[2] Y. Lin, C. Z. Vivianna Lieu, W. Zhang, L. F. Cranor, and S. Scheffler, "Measuring user responses to age verification architectures: Evidence from a deceptive online experiment," IAB/W3C Workshop on Age-Based Restrictions on Content Access (agews), 1 2025, https://datatracker.ietf.org/doc/slides-agews-paper-measuring-user-responses-to-age-verification-architectures-evidence-from-a-deceptive-online-experiment/.

[3] S. J. De and D. Le Métayer, "Priam: A privacy risk analysis methodology," in *Data Privacy Management and Security Assurance (DPM) and Quantitative Aspects in Security Assurance (QASA)*, ser. Lecture Notes in Computer Science, vol. 9963. Springer, Cham, 2016.

[4] European Commission, "Use case manuals - eu digital identity wallet," https://ec.europa.eu/digital-building-blocks/sites/spaces/EUDIGITALIDENTITYWALLET/pages/896827987/Use+case+manuals, 2025, *EU Digital Identity Wallet* resource providing descriptions of practical use cases for EUDI Wallets, including user journeys and technical/legal context.

[5] M. Frigo and a. Shelat, "The longfellow zero-knowledge scheme (draft-google-cfrg-libzk-01)," Internet Draft, IETF Crypto Forum Research Group, Sep. 2025, work in Progress; expires 6 March 2026. [Online]. Available: https://datatracker.ietf.org/doc/draft-google-cfrg-libzk/

[6] C. Paquin, G.-V. Policharla, and G. Zaverucha, "Crescent: Stronger privacy for existing credentials," Cryptology ePrint Archive, Paper 2024/2013, 2024. [Online]. Available: https://eprint.iacr.org/2024/2013

[7] E. Tweneboah, C. W. Munyendo, and Y. Zou, ""No, I can't be a security personnel on your phone"': Security and privacy threats from sharing infrastructure in rural ghana," in *34th USENIX Security Symposium, USENIX Security 2025, Seattle, WA, USA, August 13-15, 2025*, L. Bauer and G. Pellegrino, Eds. USENIX Association, 2025, pp. 5131–5148. [Online]. Available: https://www.usenix.org/conference/usenixsecurity25/presentation/tweneboah

[8] European Commission, "Commission releases enhanced second version of the age-verification blueprint," *European Commission*. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/eu-age-verification

[9] ——, "Age verification use case manual," 2026, https://ec.europa.eu/digital-building-blocks/sites/spaces/EUDIGITALIDENTITYWALLET/pages/930450954/The+Age+Verification+Manual.

[10] ——, "EU age verification solution," 2025, https://ageverification.dev/.

[11] European Parliament and Council of the European Union, "Regulation (eu) no 910/2014 of the european parliament and of the council of 23 july 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing directive 1999/93/ec (consolidated version: 18 october 2024)," https://eur-lex.europa.eu/eli/reg/2014/910/2024-10-18, 2024, consolidated text incorporating amendments up to 18 October 2024; original OJ L 257, 28.8.2014, pp. 73–114.

[12] S. J. De and D. Le Métayer, "Privacy harm analysis: A case study on smart grids," in *2016 IEEE Security and Privacy Workshops (SPW)*, 2016, pp. 58–65.

[13] S. J. De and D. L. Métayer, "A refinement approach for the reuse of privacy risk analysis results," in *Annual Privacy Forum, APF*. Springer, 2017.

[14] CyberPeace Institute, "Report of second expert meeting on harms methodology," https://cyberpeaceinstitute.org/wp-content/uploads/2024/10/Second-Expert-Meeting-Harms-Methodology-2024.docx.pdf, 2024. [Online]. Available: https://cyberpeaceinstitute.org/wp-content/uploads/2024/10/Second-Expert-Meeting-Harms-Methodology-2024.docx.pdf

[15] D. K. Citron and D. J. Solove, "Privacy harms," *Boston University Law Review*, vol. 102, p. 793, 2022.

[16] European Digital Rights (EDRi), "Online age verification and children's rights," 10 2023, https://edri.org/wp-content/uploads/2023/10/Online-age-verification-and-childrens-rights-EDRi-position-paper.pdf.

[17] J. Kelley and A. Schwartz, "Age verification mandates would undermine anonymity online," Electronic Frontier Foundation, 3 2023, https://www.eff.org/deeplinks/2023/03/age-verification-mandates-would-undermine-anonymity-online.

[18] M. Koning, P. Korenhof, G. Alpár, and J.-H. Hoepman, "The ABCs of ABCs: an analysis of attribute-based credentials in the light of data protection, privacy and identity," in *Internet, Law & Politics : A decade of transformations*, J. B. Padullés, A. C. i Martínez, M. P. Poch, I. P. López, M. J. P. de Moner, and M. V. Solana, Eds., vol. 10, no. 19. Barcelona: Huygens Editorial, 2014, pp. 357–374.

[19] L. Gak, S. Olojo, and N. Salehi, "The distressing ads that persist: Uncovering the harms of targeted weight-loss ads among users with histories of disordered eating," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW2, Nov. 2022. [Online]. Available: https://doi.org/10.1145/3555102

[20] Z. Moti, A. Senol, H. Bostani, F. J. Z. Borgesius, V. Moonsamy, A. Mathur, and G. Acar, "Targeted and troublesome: Tracking and advertising on children's websites," in *2024 IEEE Symposium on Security and Privacy*. San Francisco, CA, USA: IEEE Computer Society Press, May 19–23, 2024, pp. 1517–1535.

[21] M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, and A. Rieke, "Discrimination through optimization: How Facebook's ad delivery can lead to biased outcomes," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, Nov. 2019. [Online]. Available: https://doi.org/10.1145/3359301

[22] Y. Wu, S. Bice, W. K. Edwards, and S. Das, "The slow violence of surveillance capitalism: How online behavioral advertising harms people," in *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, ser. FAccT '23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 1826–1837. [Online]. Available: https://doi.org/10.1145/3593013.3594119

[23] J. M. Paterson, S. Chang, M. Cheong, C. Culnane, S. Dreyfus, and D. McKay, "The hidden harms of targeted advertising by algorithm and interventions from the consumer protection toolkit," *International Journal on Consumer Law and Practice*, vol. 9, pp. 1–24, 2021.

Appendix A.
DIF-based Age Verification

A.1. Model

Age verification requires that a user prove their age is above or below a certain threshold to a service provider using verified identity information issued by a trusted party such as a national government. For instance, services for which regulation requires users to be of age include adult content or gambling websites; or children-only spaces.

The European Commission proposes to implement a system for age verification that is integrated with the EUDIF [8], [9], [10]. We describe how age verification performed via the EUDIF is envisioned [9], and how we capture this formally in a functionality. The functionality is illustrated at a high level in Figure 1 This functionality abstracts from concrete cryptographic mechanisms and instead specifies trusted behavior, information flows, and enforced constraints. The regulation specifies that the EUDIF shall ensure *security by design* [11, Chapter II, Section 1, Art. 5a 12], and that "[The interoperability framework] shall facilitate the implementation of privacy and security by design." [11, Art. 12 3(c)]. To capture this requirement, we model a *minimal functionality* that reveals only the information explicitly needed to enable the intended operation of the EUDIF and does not permit any additional leakage. This minimal functionality captures an ideal privacy-by-design approach, i.e., the most privacy-preserving version of the system possible.

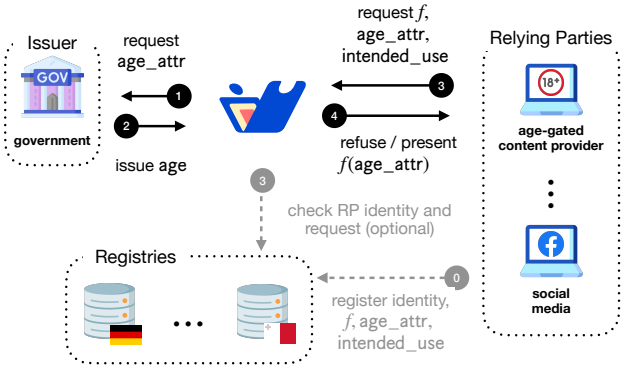


Figure 1: Overview of Age Verification for an Online Service using the EUDIF

In the functionality, we consider four types of parties: issuers $\mathcal{I}_{\mathsf{iid}}$, users $\mathcal{U}_{\mathsf{uid}}$, relying parties $\mathcal{R}_{\mathsf{rid}}$, and a content authority CA. The functionality provides interfaces for relying party registration and deregistration, attribute issuance, attribute presentation (for accessing services), and regulating content that requires verification.

First, a central entity determines which content requires age verification, for example, through regulation. In our abstraction, we model this as the *content authority* CA, who mandates that an RP $\mathcal{R}_{\mathsf{rid}}$ should register as an RP with a registration authority. RPs are not necessarily trusted (e.g., a gambling site). Our functionality describes RPs as separate entities from the issuer, although there may be scenarios where an RP is the same entity as the issuer. We model the RP's decision on whether to register with the authority as an oracle $\mathcal{O}_{\mathsf{reg}}$, capturing that an RP may opt not to implement age verification.

We parametrize the age verification functionality with a function f that captures the mandate of content authorities to verify that users are older or younger than a threshold age (modeled as an integer (age) value n). For simplicity, our functionality assumes that all RPs have the same legal requirement, i.e., all of them share the same function f. It is trivial to adapt our definition to accommodate regulations in which different RPs have different policies.

Next, a user engaging with the system may request issuance of a proof of age attestation from an issuer. The issuer for age verification is a trusted authority, such as a national or state government. To model issuance and that it is the issuer that determines attributes, we abstract the verification of real-world identity and attributes as an *issuance oracle* $\mathcal{O}_{\mathsf{Issue}}$. This models the scenario in which an issuer can verify attributes without additional information flows.

To access a service, the user is expected to prove that their age is above (or below) a pre-defined threshold, i.e., provide a proof of a binary function of a single attribute. A failure in this proof should result in a denial of access to the service. The RP requests an age credential from the user. When an RP requests data from a user, the functionality and the user interact over several rounds, during which the user is given the possibility to learn more about the request and the RP, and to decide whether to approve the request. The user's decision of whether to check the RP's registration status is abstracted into an oracle $\mathcal{O}_{\mathsf{check}}$. The user's decision of whether or not to engage with the age verification system (based on the information they are provided from the RP) is abstracted into an engagement oracle $\mathcal{O}_{\mathsf{eng}}$. These oracles model user behaviour when faced with decisions of whether or not (and how) to engage with the system. If the user has not instructed the functionality to abort, the functionality computes the output of the function f on the user's age attribute and sends it to the RP (this returns 1 if the user correctly proves that their age is above or below the pre-defined threshold).

The decision of the RP regarding whether to grant or deny access to the user, given the output of f, is captured by the oracle $\mathcal{O}_{\mathsf{ver}}$. We opt to outsource this decision to an oracle rather than directly providing access if $v = 1$ (i.e., age verification succeeds) because there is no practical mechanism to enforce RPs to provide access if verification passes: the RP obtains the output of the function f and then must decide how to act based on this verified information.

A.2. Harm trees

To better understand the origin of harms and how they arise, we develop a methodology based on harm trees [12], [13]. The primary goals of this methodology are to identify

harms that arise from each use case, and to distinguish whether a harm is inherent to the use case functionality, is rooted in the functionality being used in a digital context, or stems from concrete implementation choices.

Tree elements. We build trees from three kinds of elements: system components, capabilities, and harms, illustrated in Fig. 2.

System components. We identify *system components* that might lead to harms. These comprise information leakage due to the use of the EUDIF, the results of an entity's interaction with the system, and requirements imposed on the entities by the EUDIF. For example, system components in the age verification use case include the leakage of information about a user's age to the RP, the decision of a user to engage with the age verification system, or the requirement to have an EUDIF-enabled device to interact with the system.

Capabilities. A set of system components can combine to enable certain *capabilities* or lack thereof on the system or its entities. Examples include the capability to track users, the incapability of a provider to provide a service in a region, or the incapability of a user to access a service.

Harms. Different capabilities or a combination of capabilities can lead to harms. We build on the definition of harm by the Cyberpeace institute as "an impairment or disruption of an entity's capacity or ability to function and exist as it otherwise would have in its usual context" [14]. This includes any negative impact on individuals, groups of individuals, or the society as a whole, from the standpoint of physical, mental, or financial well-being or reputation, dignity, freedom, acceptance in society, self-actualization, domestic life, freedom of expression, or any other fundamental right. These harms can be classified within the taxonomy of privacy harms of Citron and Solove [15] and the CyberPeace Institute's harms methodology [14].

Constructing trees. When system components enable capabilities that in turn result in harms, we say that we have found a *pathway* to harm. We combine these pathways into a *harm tree* [12] that captures all pathways through which a harm may be caused. A harm tree consists of three levels. At the top of the tree is the *harm*. On the second level, we have the *capabilities* that enable this harm. At the leaves, we have *system components* that (combine in some way to) enable each capability. When representing trees, we use AND and OR operators to indicate how system components combine to create capabilities. Whenever more than one system component might be needed for the capability, we write OR^+.

Harm origins. Given a harm tree, we identify the root cause of each pathway to harm, according to the origins of its system components:

Components inherent to the functionality. In **black**, we identify components that correspond to leakage, requirements, or interactions in the EUDIF or use case functionality. Its presence in the functionality implies that these components

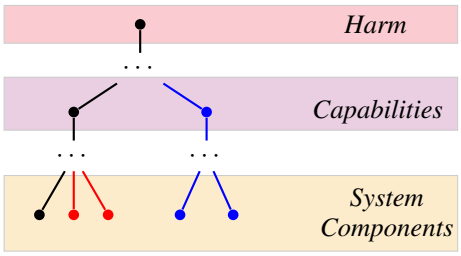


Figure 2: Harm tree. Components in black originate in the EUDIF functionality, components in blue originate in the digital context of a use case, and components in red appear due to digitalization implementation choices.

cannot be eliminated without preventing the system from operating as intended.

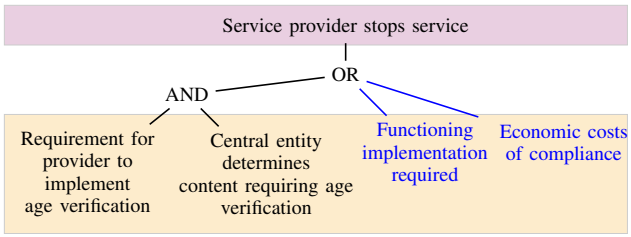
Components inherent to the digital context. In **blue**, we identify components that are inherent to the requirement that the EUDIF must be used in a digital context. They are inherent to the use of the EUDIF because they cannot be eliminated from a *digital* system.

Components linked to implementation design decisions. In **red**, we identify components that arise in a *concrete implementation* of the EUDIF. While the formal model is designed to only reveal the information necessary to realize the functionality, in reality, an implementation might leak more information. In our initial analysis, we only identify red components corresponding to a naive implementation of the use cases using the EUDIF functionality where users send their complete attributes (e.g., a copy of their identity documents) to the RP upon each presentation.

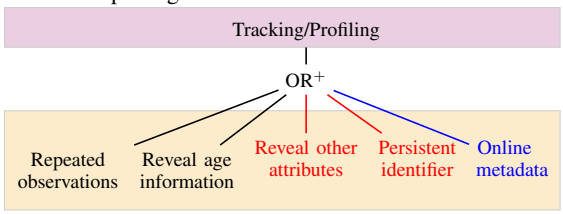
We construct harm trees for the harms that arise from the use of the EUDIF for online age verification. We extract potential harms from existing literature [16], [17], [18] and expand this set with harms discussed in [15], [14]. We find seven harms: discrimination, exacerbation of structural exclusion, targeted advertising harms, reduced availability of resources, censorship, diminished Internet safety, and increased cybersecurity risk. For system components that are inherent to the minimal functionality of the use case (i.e., are illustrated in **black**), we specify the corresponding element of the ideal functionality $\mathcal{F}_{\text{AV}}$.

Common Sub-trees In our analysis, we discover two capabilities that are common to several harms: tracking and ceasing service. For ease of exposition, we explain these sub-trees first. They are illustrated in Figure 3.

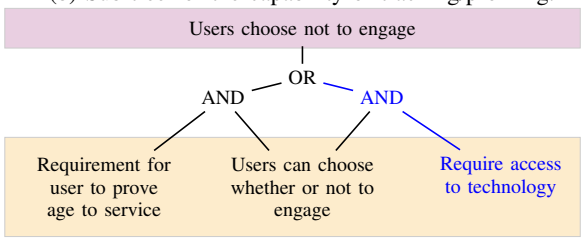
Service provider stops service. (Figure 3a). Due to implementation challenges or fear of legal consequences, service providers might decide to not implement age verification and either *stop* providing a service altogether or pull out of specific regions. Legal requirements to implement age verification are inherent to the basic functionality: A service provider receives a mandate from an authority and must register to provide a service (see **black** components). The requirement to operate in a digital context might necessitate expertise to develop a functional implementation, which



(a) Sub-tree for the capability of service providers stopping to provide service/pulling out of territories



(b) Sub-tree for the capability of tracking/profiling.



(c) Sub-tree for the capability of users choosing not to engage

Figure 3: Age verification – Common sub-trees of capabilities

might be unavailable to some service provider or incur high costs (see **blue** components).

Tracking and Profiling. (Figure 3b). Some amount of information leakage to the RP can enable service providers to perform tracking and/or profiling of the users of the system. The only information leakage to the RP that is inherent to the use case functionality is information about the user's age verification. The RP might learn additional information (e.g., online metadata or personal information) from the digital context of deployment (e.g., a browser or application). In the naive, concrete implementation where all attributes are revealed, the presence of a persistent identifier (or attributes other than age) trivially enables tracking and profiling. Many combinations of these components suffice to grant the RP tracking and profiling capabilities. In fact, even if privacy technologies are used to minimize the information revealed and eliminate persistent identifiers, the functionality inherently reveals age information that can be used for profiling.

Users choose not to engage. Requiring users to prove their age to a service provider (to access the service) or requiring them to access certain technology (e.g., devices) may result in users choosing *not* to engage with the service (Figure 3c). While the requirement for access to technology arises only due to digitalization, the requirement for users to prove their age is inherent to any system. Users may opt not to engage for fear of tracking and surveillance, due to distrust in the service provider, or because they do not want to reveal their age information. Thus, this capability arises regardless of how the system is implemented.

Discrimination (Figure 4). Discrimination harms are those that involve entrenching inequality and disadvantaging people based on their attributes (e.g., gender, race, nationality, age) [15]. Discrimination based on age information revealed during age verification is inherent to the functionality. While the denial of access to content based on age is the intended outcome, an adversary who has access to age information may provide discriminatory service (e.g., discriminatory pricing). The inherent leakage of age information, additional leakage of online metadata due to the EUDIF's digital context (e.g., cookies, fingerprinting, tracking, advertising IDs), or persistent identifiers created in a naive implementation can enable profiling capabilities which in turn might lead to discriminatory harms.

Profiling due to leakage of persistent identifiers can be mitigated by privacy-preserving solutions that provide unlinkability. However, even privacy-preserving solution cannot prevent profiling based on age information and information from the digital context.

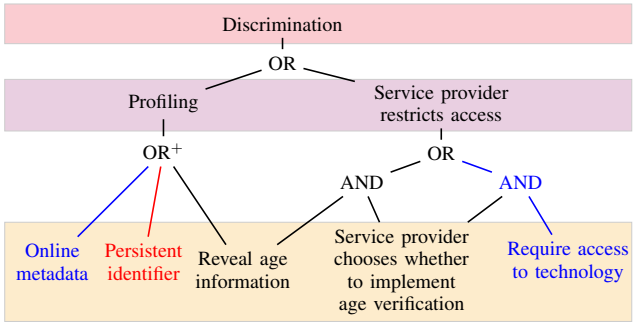


Figure 4: Age verification – Discrimination

Exacerbate structural exclusion/restricted access to resources (Figure 5). Structural exclusion restricts individuals' choices and freedom of information, and can be considered an *autonomy harm* [15]. In the context of age verification, exclusion can happen due to two capabilities. First, users may not be able to access to the service. This affects children that do not fulfill the age requirements, but also adults that cannot verify their age (e.g., those with low digital literacy, or with no access to official identity documents). Second, service providers might pull out of territories where the market size does not offset the added costs of complying with age verification regulations, which reduces access to services for users from certain geographical regions.

Targeted advertising harms. Targeted advertising has been shown to lead to psychological harm [19], [20], result in discriminatory practices [21], and restrict users' opportunities for choice and exercises control over their actions [22], [23]. Use of the EUDIF for online age verification is set to further deepen these harms. Minors can be targeted for

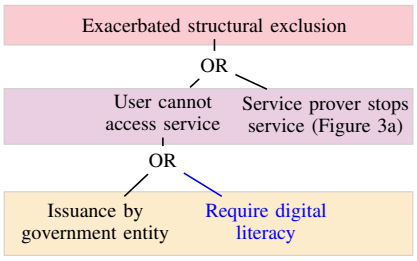


Figure 5: Age verification – Structural exclusion

advertisements based on verified age information that must be revealed to achieve the functionality of the system. Adults can be targeted based on their verified age and additional information leaked through the digital context directly or via data brokerage. A naive implementation increases the risks of targeting, by enabling better tracking and profiling (see Fig. 3b). Even if privacy enhancing technologies are used to limit profiling, the risk of targeted advertising harms persist due to the age-related leakage.

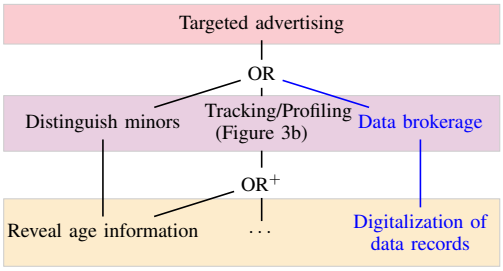


Figure 6: Age verification – Targeted advertising

Reduced availability of resources The introduction of the EUDIF might lead to a chilling effect on legitimate activities. As previously noted, less information might be published because of the additional burden on operators to implement age verification. Individuals may refrain from publishing content online if must proof age to publish. Service providers might pull out of territories due to additional burden of compliance which means fewer services might be available in one region. The paths to this harm are inherent to any system implementing age verification for online services, illustrated in Figure 7.

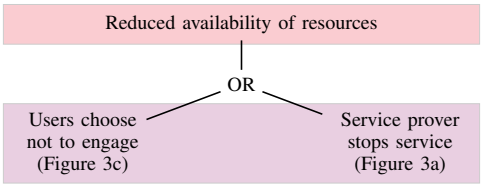


Figure 7: Age verification – Reduced availability

Censorship The EUDIF system, when used for age verification, can lead to censorship in multiple ways. A central entity is given the power to decide which services require age verification. This central entity can restrict what children can see online and which platforms and services they can use. Thus ultimately impairs children's self-development (due to biased educational content) and restricts their freedom of expression (by blocking access to private messengers or social media). Furthermore, function creep/repurposing of the system is enabled by the central entity deciding which content requires age verification or low-integrity operators/service providers choosing whether or not to implement age verification.

The EUDIF system also leads to some forms of self-censorship. Individuals might refrain from publishing content online if it requires using the system to prove their age, undermining freedom of expression. The resulting censorship harm (either via self-censorship, or imposed by the operators or central entity) is inherent to the functionality of the use case.

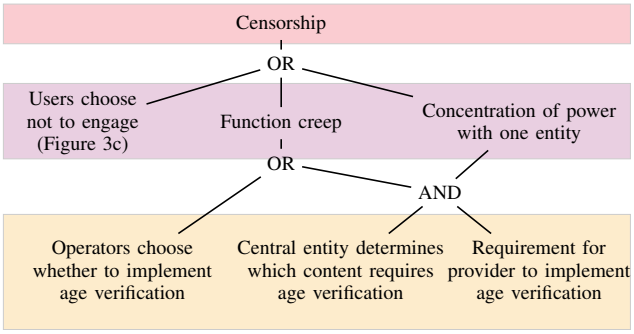


Figure 8: Age verification – Censorship

Diminished Internet safety The EUDIF bears potential for tracking and de-anonymising online users and relies on an official authority to issue a proof of age to to access certain services. As a result, it prevents anyone who needs independence from the issuing authority or anonymity online from safely using the internet. Certain groups, such as investigative journalists, activists, or some researchers are prevented from doing their jobs. Users are discouraged from expressing unpopular opinions if they need to fear de-anonymisation. This is harmful because it erodes democratic freedoms such as freedom of expression and opinion.

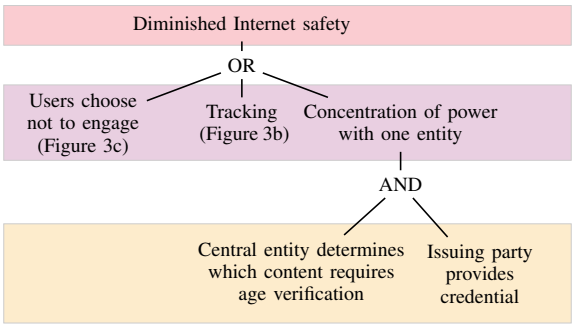


Figure 9: Age verification – Diminished Internet safety

Increased cybersecurity risks For similar reasons that an operator may opt to stop providing its service, it may opt to simply become non-compliant. That is, we can replace the capability in Figure 3a with "Service provider becomes unregulated." Additionally, we have already established that users may opt not to engage with the age verification system (Figure 3c). This suggests that users might be pushed to unsafe/unregulated spaces and the amount of such unregulated services will increase. A naive implementation exacerbates this harm even further, by normalizing requests for, and increasing retention of, identity information. Altogether, this increases users' exposure to risks such as fraud or theft.

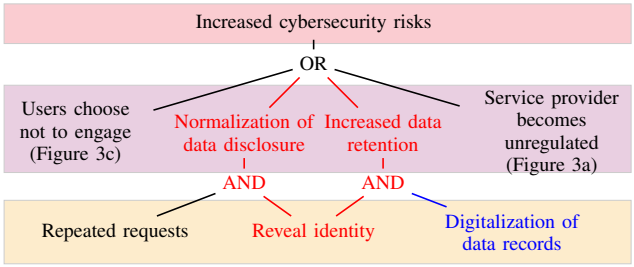


Figure 10: Age verification – Increased cybersecurity risks