JAPAN IGF 勉強会

子どもを守るために すべてを見張る？

オンライン年齢認証・チャット検閲と、暗号化の終わり

八田 真行

駿河台大学 経済経営学部 ／ 日本IGF運営委員

2026年7月24日

AGENDA

本日の流れ

7つの角度から「オンライン児童保護」規制の技術的実態を検証します。

01 **世界の規制の潮流**

豪・英の年齢制限、EUのChat Controlと「年齢認証アプリ」

02 **「ネットは子どもに悪い」説の出所**

Haidt、ドラマ『アドレセンス』と、モラルパニックの構図

03 **CSAM誤検出の技術的実態**

検知アルゴリズムの精度は本当に主張通りか

04 **年齢認証義務化の代償**

顔認証の限界と、大人を含む全利用者のプライバシー

05 **暗号を弱めることの危険性**

Keys Under Doormats・Dual_EC_DRBG・グルーミング論への反論

06 **代替案① 手元で完結する年齢認証**

アプリストア／端末レベルでの年齢証明という設計

07 **代替案② プロトコル・ベースの規制**

「すべてを見張る」ことなく秩序をつくる、もう一つの道

INTRODUCTION

世界中で同時多発する規制の議論

2024年末以降、18歳未満のSNS利用制限を検討・可決・施行する動きは、豪州を皮切りに英国・EU加盟各国・インドネシア・マレーシア・トルコなど、世界各地で同時多発的に広がっています。「子どもを守る」という大義名分には正面から反対しにくく、規制は各国で異例の速さで進んでいます。

実際に組み込まれようとしている設計は、驚くほど共通しています。AIによる自動判定（年齢推定とコンテンツ検知）を通信インフラそのものに組み込み、国家が個人の通信内容や属性の審査に構造的に介入する、という仕組みです。

年齢を推定するために顔と行動を分析し、危険を検知するためにメッセージを走査する。「保護」の名の下に監視されるのは、子どもだけではなく国民全員です。

この仕組みは、本当に「子どもを守る」という約束を果たせるのか——これが本日の問いです。

出典：TechCrunch（2026/7/21）、CEPA "Mapping the Spread of Child Safety Rules"（2026）

1

PART 1

世界の規制の潮流

豪州・英国のSNS年齢制限、EUの「Chat Control」と年齢認証アプリ。手法は違っても、向かう先は同じです。

豪州 — 世界初「16歳未満SNS禁止」の制度設計

2025年12月10日、16歳未満のSNS利用を法律で禁じる世界初の制度（Online Safety Amendment: Social Media Minimum Age Act 2024）が施行されました。対象はFacebook、Instagram、Snapchat、TikTok、YouTube、X、Reddit、Twitch、Threads、Kickなど10のプラットフォームです。

罰則の対象はあくまで事業者であり、子どもや保護者本人が処罰されることはありません。制裁金は最大9,900万豪ドルで、2026年6月にはその倍増も示唆されています。

2025/12/10
施行日

10
対象プラットフォーム数

最大A$9,900万
制裁金（2026/6に倍増示唆）

出典：eSafety Commissioner、The Guardian（2026/6/27）

豪州 — 数字が語る「形骸化」

470万件

2026年1月時点で削除・制限された、16歳未満とみなされたアカウント

約70%

禁止後も引き続きSNSにアクセスしていたと回答した16歳未満（保護者調査、eSafety）

しかし施行からわずか数か月で、形骸化を示す数字が並びました。保護者調査では16歳未満の約70%が禁止後もSNSへのアクセスを続けたと回答しています。顔ベースの年齢推定はマスクや年長の兄姉の「代打」で突破される事例が多数報告され、行動ベースの推定には1〜3歳の誤差が伴います。

2026年3月31日、eSafetyはFacebook、Instagram、Snapchat、TikTok、YouTubeの5社に対し、不遵守の疑いで正式調査を開始しました。施行前から指摘されていた技術的限界をそのままなぞる形になっています。

出典：eSafety Commissioner 遵守レポート（2026/3）、The Conversation（2026）

英国 — Online Safety Actと「豪州プラス」への布石

英国では2025年7月25日、Online Safety Act（OSA）の児童保護義務が本格施行されました。ポルノや自傷・摂食障害コンテンツ等には「highly effective age assurance（HEAA）」と呼ばれる高精度の年齢確認が要求され、制裁金は最大1,800万ポンドまたは世界売上高の10%に達します。実際、RedditにはID年齢確認義務違反等で1,447万ポンドの制裁金が科されました。

それでも、13歳未満は利用できないはずのSNSに8〜12歳の72%がなおアクセスしていたという調査があります。さらに2026年6月、スターマー首相（当時）は16歳未満の全面SNS禁止の方針を表明し、16〜17歳向けの「ソーシャルメディア門限」の導入も同時に浮上しました。英国は「豪州プラス」へと踏み込みつつあります。

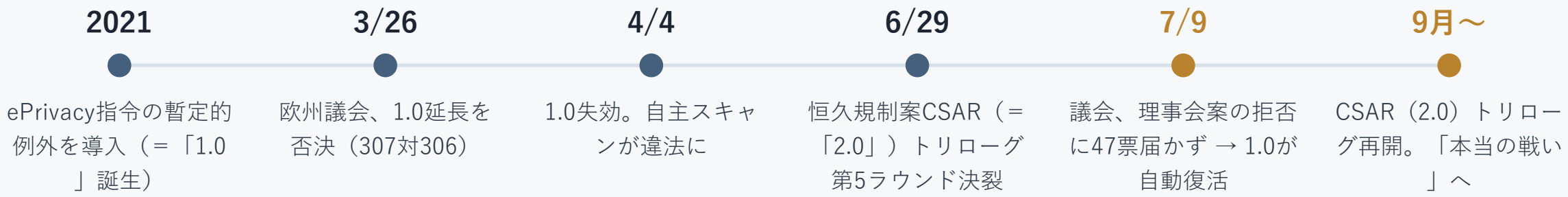
72%
13歳未満限定のはずが、なおアクセスしていた8〜12歳の割合

£1,447万
Redditに科された制裁金（年齢確認義務違反等）

約7万件
Discordの年齢確認委託先から流出したID画像（Part 4で詳述）

出典：Ofcom、ICO、gov.uk、TechRadar（2026/6）

EU「Chat Control 1.0」— 2026年、綱引きの半年



EUでは「Chat Control」と呼ばれる、児童保護を名目とした通信内容スキャン施策がずっとくすぶっていて、迷走しています。最近の動向で鍵になるのは議会の第二読解ルールです。理事会案の「拒否」には議員総数の絶対多数（720議席中361票）が必要で、棄権や欠席は事実上、理事会案への賛成として扱われます。2026年7月9日の採決では拒否が47票届かず、いったん失効していた「1.0」が自動的に復活しました。

「1.0」と「2.0」は別物 — 混同すると論点を見失う

Chat Control 1.0

暫定・任意の「例外規定」です。GmailやMessenger、Skype、Instagram DM（2026年5月以降）といった非暗号化サービスを対象に、企業が「自主的に」スキャンする建付けを取ります。2028年4月まで延命されましたが、E2EEサービスを対象外とする修正が7月9日に採択されました。

CSAR（Chat Control 2.0）

恒久・義務的な「規則案」です。司法・当局の「検出命令」により、特定サービスにスキャンを義務付けます。最大の争点はE2EEサービスにもクライアントサイド・スキャン——暗号化される前に端末上でAIが中身を検査する仕組み——を義務化するか否かで、トリローグは5回連続で決裂し、9月に第6ラウンドが予定されています。

最大20%

欧州委員会自身の実施報告書が示す、未知素材を検知するAI分類器の誤検知率。フラグが立った通信の5件に1件が無関係という水準です。

EU ― 「年齢認証アプリ」と2026年7月以降の展開

2026年4月、欧州委員会はEU共通の「年齢認証アプリ」を発表しました。フォン・デア・ライエン委員長はこれを「世界最高水準のプライバシー基準」を備え、完全にオープンソースで、どの端末でも動作すると説明していました。プラットフォームにアプリの利用義務はありませんが、「同等に有効な」年齢確認を示せなければ、DSA（デジタルサービス法）上の制裁の対象になります。

続く7月13日、委員長は「子どもの安全に関する特別パネル」の報告書を受け取り、夏以降に、SNSの最低利用年齢（「ソーシャルメディア・スタート・エイジ」）を含む法案を提案すると表明しました。飲酒や自動車運転の年齢制限に触れつつ、シートベルトの定着と同じ「時間のかかる文化的変化」だと位置づけます。一方、アプリが「万能ではない（It won’t be foolproof）」ことも自ら認めています。

「これは子どもがソーシャルメディアにアクセスできるかの問題ではない。ソーシャルメディアが、いつから私たちの子どもにアクセスできるかの問題だ」― フォン・デア・ライエン（2026/7/13）

出典：European Commission、TIME（2026/4/16）、ID Tech（2026/7/13）、Euronews（2026/7/13）

「児童保護はトロイの木馬」— 実務家からの根本的批判

30年にわたりオンライン児童保護技術に携わってきたPaul Walsh（AOLでAIMの立ち上げに参画し、W3Cのコンテンツラベリング標準を共同創設、NCMEC（全米行方不明・被搾取児童センター）とのMOUの下で検知技術を構築してきた人物）は、委員長の発言を逐条的に批判し、EUの児童保護レトリックを「完全なデジタル監視と統制のためのトロイの木馬」だと断じました。

「プライバシー保護型の年齢認証」は語義矛盾

年齢は本人確認なしには検証できず、確認後にどう共有されるかは本質ではない、と指摘します。アプリは公開直後に破られたとも述べています。

車の鍵や酒類販売との比喩は誤り

免許や身分証の確認を「全員に・あらゆる入口で」求める制度は現実には存在せず、比喩は監視の正当化にすぎないと論じます。むしろ親の権限を国家が奪う設計だ、というのがWalshの見立てです。

「万能ではない」という自認こそ核心

回避手段（VPN）を塞ごうとすれば、VPNの利用自体を本人確認制にするほかなく、規制は必然的に拡大する、と警告します。

※ 欧州委員会側は「インターネットの検閲ではない」と反論しています（Euronews, 2026/7/13）。ただし「作る側」を熟知した専門家による内在的批判として、検討に値します。

出典：Paul Walsh, LinkedIn（2026/7/13）

手法は違えど、設計思想は共通している

AIによる自動判定の実装

豪州・英国は年齢推定を、EUはコンテンツ検知を、通信・OSインフラそのものに組み込もうとしています。判定の主体は人間からアルゴリズムへ移ります。監視のためには暗号化された通信経路を弱めるのも辞さない方向です。

「児童保護」という不可侵の大義名分

実効性の検証よりも、義務化と処罰規定の整備が先行します。反対論は「子どもを守りたくないのか」という問いの前で萎縮させられます。

国家による属性・内容審査の構造化

個々のサイトではなく、通信経路・端末レベルでの介入が標準化されつつあります。年齢認証アプリは、その「入口」を全員に義務づける仕組みです。

次の問い：この設計は、本当に「約束」を果たせるのか？　まず、規制を支える科学的根拠から検証します。

2

PART 2

「ネットは子どもに悪い」説の出所

Jonathan Haidtのベストセラーと、Netflixドラマ『アドレセンス』。規制を駆動する「通説」は、どのように作られたのでしょうか。

規制を支える一冊の本 — 『The Anxious Generation』

規制論の理論的支柱となっているのが、Jonathan Haidt『The Anxious Generation』（2024、邦訳『不安の世代』）です。スマートフォンとSNSの普及——著者のいう「The Great Rewiring（大配線替え）」——こそが、2010年代に世界の若年層を襲ったメンタルヘルス危機の主因である、という主張は世界的ベストセラーになりました。

本書は豪州のSNS禁止、英国OSAの強化、EUの児童保護論議など、各国の規制論議で理論的支柱として繰り返し引用されています。著者自身の発信力とメディア露出も大きく、政策担当者や保護者の間では、もはや検証を要しない「通説」と化しています。

「因果関係」は示されているか — 専門家の反論

しかし、科学的合意は存在しません。Candice Odgers（UCアーバインの心理学者）は『Nature』誌の書評でこう総括しています。「私を含む数百人の研究者が、Haidtが主張するような大きな効果を探してきた。得られたのは『効果なし・小さい・混在』という結果ばかりだった。データの多くは相関関係であり、因果関係ではない」。

スクリーンタイムと幸福度の大規模計量研究を主導してきたAmy Orben（ケンブリッジ大）とAndrew Przybylski（オックスフォード大）も、先行研究（Twenge等）の再現を試みた結果、効果は小さく不安定だと指摘します。『Science』誌編集長H. Holden Thorpの「これだけの影響力があるなら、科学的不確実性を伝える責任があるのでは」という問いかけに対する、Haidtの答えは「そうは思わない」でした。

出典：Odgers, C.L. "The great rewiring", Nature 628 (2024)；Platformer (2024)

根拠と政策対象の「ズレ」

実験対象と政策対象の不一致

Haidtが因果の根拠とする「数十の実験」は、そのほぼ全てが大学生や18〜35歳の成人を対象としており、13歳未満を対象にした実験は一件もありません。しかし規制が対象とするのは、まさに児童・若年ティーンそのものです。

統計そのものの読み方の問題

自傷による救急外来受診の増加も、2011年のスクリーニング基準変更と2015年の傷害の「意図性」コーディング変更という、測定方法自体の変化が絡んでおり、単純な「悪化」として読むことはできません。

つまり、「若者の不調が増えた」という観察と、「その原因はSNSであり、禁止すれば改善する」という政策命題の間には、いくつもの飛躍が挟まっているのです。

もう一つの起爆剤 — Netflixドラマ『アドレセンス』（2025）

2025年3月に配信された英国発のドラマ『アドレセンス（Adolescence）』は、13歳の少年が同級生の少女を殺害した事件を、全4話・各話ワンカット撮影で描き、Netflix史上有数のヒットとなりました。少年がオンラインの「マノスフィア」（ミソジニー的な男性圏コミュニティ）やインセル文化に影響されていたという筋立てが、「SNSが子どもを蝕む」という社会の不安に、強烈な物語的実感を与えたのです。

反響は政治を直撃しました。スターマー首相（当時）は制作者らを首相官邸に招いて政策議論を行い、Netflixは英国の全中等学校に本作を無償公開しました。フィクションであるにもかかわらず、本作は議会審議やスマートフォン規制論のなかで、しばしば「証拠」であるかのように語られ、2026年のSNS禁止方針への世論的な地ならしになったと評されています。

一本のドラマが政策論議の「エビデンス」として機能することへの懸念は、研究者や批評家から繰り返し示されましたが、熱狂の中でかき消されがちでした。「エビデンスではなくアネクドート」はポピュリズム政治の定番です。

「モラルパニック」— 半世紀繰り返されてきた構図

社会学者Stanley Cohenは『Folk Devils and Moral Panics』（1972）で、ある集団や事物が「社会の価値を脅かす存在」としてメディアにより誇張され、世論と当局の過剰反応を引き起こす過程を「モラルパニック」と名付けました。脅威の誇張、単純化された悪玉（folk devil）の指名、専門家の警告、そして拙速な立法——という定型的な経過をたどります。

新しいメディアは、いつもその標的でした。1950年代の米国ではコミックが少年非行の元凶とされ（『無垢への誘惑』とコミックス・コード）、その後もロック音楽、テレビ、テレビゲーム（暴力ゲーム規制論）、日本の「ゲーム脳」「スマホ脳」論などが同じ道をたどりました。いずれも当時は因果関係が自明とされましたが、後の研究で、主張された効果は確認されないか、ごく小さいことが分かっています。

ベストセラー（Haidt）が「科学」の顔を、ドラマ（『アドレセンス』）が「物語」の顔を与え、両者が立法を駆動する——現在のSNS規制論は、この古典的な構図の最新版として読むことができます。

「悪いコンテンツの存在」と「行動の変容」は別の話

「有害なコンテンツがSNS上に大量に存在する」という観察と、「そのコンテンツが実際に子どもの行動や思想を変えている」という主張の間には、証明を要する大きな飛躍があります。過去の「メディアが子どもを壊す」論の多くは、この二つを暗黙のうちに同一視してきました。暴力表現を含むテレビゲームが典型例です。1990年代以降、数百本規模の研究が積み重ねられましたが、Christopher Fergusonらの系統的なメタ分析は、対人暴力への因果的効果は事実上ゼロに近いと繰り返し示しています。「悪いコンテンツ」は昔から存在し続けてきましたが、それが行動を変えたという主張のほうは、検証のたびに縮小してきたのです。

なぜ両者はしばしば混同されるのか。有害コンテンツへの実際の接触は、少数の強い動機を持つ層に集中する傾向があることが、大規模な実証研究の総括で示されています。人は受動的に「落ちる」のではなく、自ら求めて踏み込むのです。加えて、「情報操作に弱いのは自分ではなく“他の誰か”だ」という第三者効果の認知バイアスが、実際の影響を過大に見積もらせることも知られています。「見えるものが増えた」ことと「変えられた人が増えた」ことは、別の主張として個別に検証されなければなりません。

出典：Ferguson, Perspect Psychol Sci 10 (2015)；Budak et al., Nature (2024)；Altay & Acerbi (2023)

子どもだけでなく「社会全体への影響」論も、根拠は同様に弱い

「SNSが分極化や陰謀論、民主主義の機能不全を引き起こした」という、子ども向け議論の背後にあるもう一段大きな通説にも、同様の弱さが指摘されています。Meta社と学術チームによる2020年米大統領選をめぐる大規模ランダム化実験では、アルゴリズム順のフィードを時系列順に切り替えても、再シェア表示を止めても、政治的な態度や分極化には検出可能な変化が生じませんでした。19,857人のFacebookアカウントを選挙前6週間停止させた実験でも、結果は「ほぼゼロ」でした。これらは因果推論の「ゴールドスタンダード」とされる手法による知見です。

分極化そのものも、SNS普及より何十年も前から進行していた長期トレンドであり、2017年の研究では1996〜2016年の分極化の伸びはSNSを最も使わない高齢層で最大でした。12カ国を比較した研究でも、同程度にSNSが普及した国々の間で分極化は逆方向に動いています。陰謀論の蔓延度についても、SNS時代に体系的な増加を示す証拠は見当たりません。政治的無知や陰謀論、エリート主導の虚偽情報は、いずれもSNS以前から社会に存在してきた病理なのです。

出典：Nyhan et al., Nature (2023)；Guess et al., Science (2023)；Boxell et al., PNAS (2017)

「子どもは特に脆弱」という前提も、思われているほど強固ではない

「子どもの脳は発達途上にあり、大人よりも影響を受けやすい」という直感は根強いものですが、これも実証データの裏づけは薄いのが実情です。Amy OrbenとAndrew Przybylskiが英国の大規模コホートデータを2万通り以上のモデルで分析した結果、デジタル技術利用が説明する幸福度の分散はわずか0.4%にとどまり、その負の関連の大きさは「ポテトチップスを定期的に食べること」や「眼鏡をかけていること」が幸福度に持つ関連とほぼ同水準でした。同じ著者らが2021年に行った追跡調査でも、この関連が年々強まっている、つまり「以前より有害になっている」ことを示す証拠は見つかっていません。

「思春期の脳は前頭前野が未成熟でリスクに無防備だ」という神経科学の通説自体にも、専門家の間で異論があります。青年期の脳の変化は、感受性を一律に高める「無防備さ」ではなく、新奇な経験を積極的に求める発達上の機能として捉えるべきだとする研究があり、単純な「未熟だから脆弱」という図式は臨床・神経科学の知見と食い違うとの指摘もあります。子どもを特別扱いする規制の前提自体、思われているほど盤石ではありません。

出典：Orben & Przybylski, Nat Hum Behav (2019)；Vuorre et al. (2021)；Bernheim et al. (2013)

この部のまとめ：弱い根拠の上に強力な監視インフラ

相関は存在しますし、「ゼロリスク」だと主張したいわけでもありません。殺人事件など個々の被害は現実であり、対策そのものは必要です。しかし、現状の規制の規模と強制力を正当化できるほど強固な因果的根拠はない、というのが研究者の間の現時点での共通理解だと思います。

にもかかわらず、ベストセラーとドラマに支えられた「通説」が、世界中の一律規制の理論的支柱になっています。

モラルパニックの定型どおり、弱い因果的根拠の上に、AIによる自動判定という強力な監視インフラが築かれようとしている——これがPart 1で見た規制の「地盤」です。

3

PART 3

CSAM誤検出の技術的実態

検知アルゴリズムの精度は、本当に公式発表の通りなのでしょうか。2026年、初めて独立検証のメスが入りました。

PhotoDNAと「500億分の1」という主張

CSAM（児童性虐待素材）検知の事実上の標準が、Microsoftが2009年に開発したPhotoDNAです。画像の「見た目」を圧縮した指紋（知覚ハッシュ）を作り、既知のCSAMデータベースと照合する仕組みで、GoogleやMeta、TikTok、NCMECなど世界中で採用され、1日に数十億枚単位の画像が照合されています。

1 / 500億

公式に主張される誤検知率。しかしこの数値の出所は、開発に関わったHany Faridによる2019年の報告のみ。算出方法は非公開で、独立した第三者による検証は一度も行われてきませんでした。

「一度も独立検証されていない数字」が、全市民の通信をスキャンする制度の安全性の根拠として、各国の立法論議で引用され続けてきた——これがまず押さえるべき出発点です。

出典：Deryck, Leblanc-Albarel, Preneel, "White-Box Attacks on PhotoDNA", IACR ePrint 2026/486

KU Leuven（COSIC）チームの発見（2026年）

2026年、KU LeuvenのMaxime Deryck、Diane Leblanc-Albarel、Bart Preneelは、PhotoDNAの数式的な挙動を世界で初めて完全に再現しました。ハッシュ値は各画素ブロックのRGB値の合計にしか依存せず、区分線形かつ微分可能――つまり勾配ベースの最適化が使え、ノートPC上で数秒から数分で、ハッシュの衝突や準衝突を生成できることが示されたのです。

検知の回避

本物のCSAM画像を、見た目をほとんど変えることなく、既知ハッシュのデータベースから外れるように加工できます。

誤検知の意図的生成

無害な画像に、既知の違法画像と同一のハッシュを持たせられます。標的に送りつければ、冤罪を機械的に作り出せることになります。

Appleが2021年に提案したクライアントサイド・スキャン用のNeuralHash（撤回後も端末上に残存）でも、同様の衝突が実証されています。チームは2025年9月にMicrosoftへ責任ある開示を行い、研究はUSENIX Security ’26に採択されました。

出典：COSIC (KU Leuven)、pseudodna.eu、IACR ePrint 2026/486、USENIX Security ’26

「規模」が誤検知を破滅的にする

誤検知は「規模」によって破滅的になります。欧州委員会自身の実施報告書は、未知素材を検知するAI分類器の誤検知率を最大20%と示しています。そして仮に精度99%という楽観的な数字を置いても、大手メッセージングサービスの規模でスキャンすれば、結果は次のようになります。

1,000億通
大手メッセージングサービスの1日のやり取り数（一例）

×

× 1%
仮に精度99%だとしても残る誤り

=

= 10億通
1日で生じる誤検知の件数

誤検知の奔流は捜査機関の処理能力を圧迫し、本物の被害児童を救うためのリソースを奪います。「脆弱で不透明な検知システムを大規模展開することは、プライバシー侵害と検知失敗の両方のリスクを高める」というのがCOSICチームの警告です。

4

PART 4

年齢認証義務化の代償

「はい／いいえ」に答えるだけのはずが、実際に集められているのは何か。そして顔認証は、そもそも年齢を言い当てられるのか。

「はい／いいえ」に答えるだけのはずが

年齢認証は建前上、「18歳以上ですか」という問いに答えるだけの仕組みのはずです。しかし実際にサイトごと・サービスごとに繰り返し集められているのは、次のようなデータです。

- **政府発行ID画像**　運転免許証やパスポート等のスキャン画像が、訪問先ごとに提出されます。
- **顔写真・動画（生体情報）**　「年齢推定」「生体照合」を目的とするライブ撮影が求められます。
- **行動データ**　検索履歴や視聴カテゴリといった、既にアカウントに紐づく信号から年齢が推定されます。
- **決済・口座情報**　クレジットカード確認やオープンバンキング連携が「年齢の証明」に流用されます。

Googleの行動ベース年齢推定は、大学受験を調べる17歳と、リスキリングで社会人入学を調べる40代を、境界線付近でどう区別するのでしょうか。技術的な説明はありません。

顔認証は、16歳と17歳を見分けられない

顔画像からの年齢推定は近年大きく進歩しました。それでも米国NIST（国立標準技術研究所）の評価（FATE Age Estimation）では、最良水準のアルゴリズムでも平均誤差はなお2〜3年程度あり、業界最高水準を掲げるYotiでさえ、13〜17歳について1年強の平均誤差を公表しています。誕生日を挟んだだけの16歳と17歳の「顔」を区別することは、原理的に不可能なのです。

このため実運用では「21歳に見えなければID提出」といったバッファ（緩衝帯）を設けざるを得ず、境界年齢の前後にいる大量の若年成人が、顔スキャンに加えて身分証の提出まで求められます。しかも誤差は性別や肌の色によって偏り、推定の失敗が特定の集団に集中するという公平性の問題も、NISTの評価で繰り返し確認されています。

±2〜3歳

最良水準のアルゴリズムでも残る、顔からの年齢推定の平均誤差のレンジ。法律が引く「16歳」「18歳」という1年刻みの線と、技術の解像度は根本的に噛み合っていません。

出典：NIST FATE Age Estimation & Verification（2024–）、Yoti White Paper

精度を「補う」ために、監視が広がる

顔で分からなければ、プラットフォームは他のシグナルで補おうとします。Metaは、友人ネットワークの年齢構成、「16歳おめでとう」といった誕生日投稿、アカウントの利用パターンなどから年齢を推定していると説明しており、Googleも検索内容や視聴動画カテゴリを使った行動ベースの推定を導入しています。

つまり、年齢推定の精度不足を埋める「解決策」とは、ソーシャルグラフ・投稿内容・行動履歴のより深い分析——すなわち、より多くのデータ収集にほかなりません。子どもを守るための仕組みが、子どもを含む全利用者の交友関係と行動を常時分析する監視エンジンへと転化する。ここに構造的な倒錯があります。そして、データ収集の対象は子どもだけではありません。

「精度が足りない」という批判が、「もっとデータを集める」ことを正当化してしまう。年齢認証の義務化は、この悪循環を法律で駆動する仕組みになりかねません。

ケーススタディ：Discordの年齢確認基盤流出

2025年10月、Discordの年齢確認の異議申し立てを処理する委託先が侵害されました。担当者1名のアカウントが破られてから58時間にわたり、攻撃者はシステムへのアクセスを維持していました。公式発表では約7万件の政府発行ID画像が流出し、攻撃側は最大210万件を主張しました（Discordは否定しています）。

その4か月後の2026年2月、Discordは「全ユーザーに顔スキャンまたは身分証の提出」を義務化すると発表し、猛反発を受けてロールアウトを延期しました。流出の教訓が「もっと集める」方向の決定につながった、という点も示唆的です。

教訓：委託先ベンダーを変えても、「身分証明書を一箇所に集める」というアーキテクチャ自体が標的であり続けます。

出典：The Record、Proton、stateofsurveillance.org（2025–2026）

誰が排除され、何が萎縮するのか

1,500万人
運転免許証を持たない米国成人

2,600万人
政府発行の写真付きIDを一切持たない米国人

3,450万人
現住所と一致するIDを持たない米国人

ID要求は、まず「IDを持たない人々」を排除します。影響は貧困層・高齢者・障害者・人種的マイノリティに偏って及びます。またジャーナリストや内部告発者にとって、匿名性の喪失は情報源特定のリスクに直結し、報道機関への政治的圧力の新たな経路になりうるとの指摘もあります。

そしてプライバシーの本質は「利用の有無」ではなく「記録の存在そのもの」にあります。成人向けサイトの年齢確認データベースに名前があるという事実自体が、閲覧内容についての推測を可能にしてしまうのです。

出典：EFF、The Intercept（2026/6/28）、Bellovin (2025) "Privacy-Preserving Age Verification—and Its Limitations"

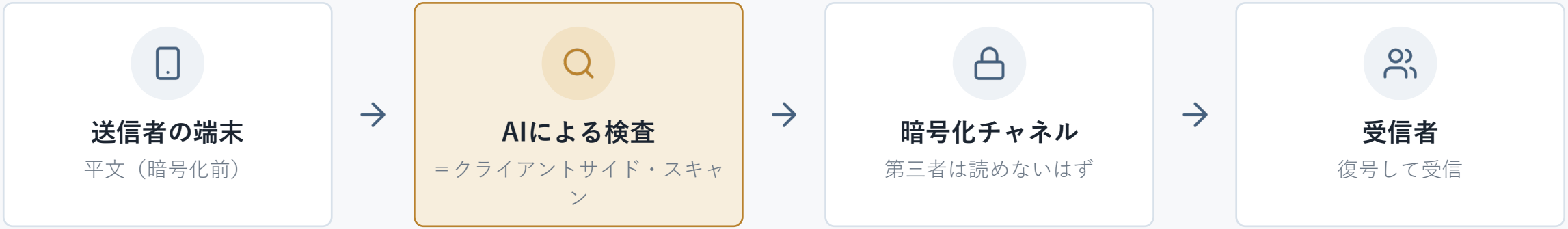
5

PART 5

暗号を弱めることの危険性

「良い人だけが通れる裏口」は作れるのでしょうか。暗号学者の40年の答えと、実際に起きた事件を見ていきます。

「暗号化されているが、検査はできる」は成立しない



E2EE（エンドツーエンド暗号化）の設計原理は、暗号鍵を送信者と受信者だけが持ち、通信経路上の誰にも内容を読ませない、というものです。クライアントサイド・スキャンは「暗号化される前に、端末上でAIが中身を検査する」という発想であり、暗号化そのものを迂回して「暗号化されているが、政府には見える」状態を作り出します。

検査対象を「CSAMだけ」に限定する技術的な保証は存在しません。スキャンの照合リストを差し替えるだけで、監視対象は際限なく拡張しえます。

暗号学者たちの警告 — 「ドアマットの下の鍵」

1990年代の「暗号戦争」（クリッパーチップ論争）を戦った世界最高峰の暗号学者・セキュリティ研究者15名——Abelson、Rivest、Diffie、Schneier、Bellovinら——は、共著論文「Keys Under Doormats」（2015）で、法執行のための「例外的アクセス」がなぜ成立しないかを体系的に示しました。論点は三つに整理されます。

ベストプラクティスへの逆行

今日の通信の安全は、鍵を使い捨てる前方秘匿性（forward secrecy）などの設計に支えられています。「後から復号できる鍵」の温存は、この土台と両立しません。

複雑性の爆発

例外的アクセスの仕組みはシステムを複雑にし、その複雑さ自体が新たな脆弱性の源になります。セキュリティ工学の鉄則は「単純さこそ安全」です。

価値の集中

全員分の通信を開けられる鍵の保管庫は、犯罪者や敵対国にとって最高価値の標的——まさに「ドアマットの下に隠した合鍵」になります。

インターネットが社会基盤そのものとなった今日、20年前のクリッパーチップ論争で示された結論はいっそう重い——これが論文の結論です。この警告が机上のものでないことは、次の事件が示しています。

出典：Abelson et al., "Keys Under Doormats" (2015), Journal of Cybersecurity

事例：Dual_EC_DRBG — 裏口は誰のものになるか

2006年、NSAが設計を主導した擬似乱数生成器Dual_EC_DRBGが、米NISTの標準（SP 800-90A）に採用されました。翌2007年にはMicrosoftの研究者らが、設計定数の選び方次第で出力を予測できる「トラップドア」を仕込める構造だと指摘します。2013年、スノーデン文書が標準化過程へのNSAの介入を裏づけ、ロイターは、NSAがRSA Securityに1,000万ドルを支払って同社の暗号ライブラリBSAFEの既定乱数生成器に採用させていたと報じました。NISTは2014年、このアルゴリズムを標準から撤回します。

決定的なのは2015年のJuniper事件です。同社のVPN製品ScreenOSで、何者かがDual_EC_DRBGの定数を「自分の鍵」に差し替える改竄を行っていたことが発覚しました（後の報道は国家系ハッカー集団の関与を指摘しています）。政府機関や企業のVPN通信は、誰とも知れぬ相手に復号可能な状態に置かれていたのです。

「良い人」のために仕込まれた裏口は、設計者の意図を離れ、より悪い誰かのものになる——Dual_EC_DRBGは、その実例です。

出典：Shumow & Ferguson (2007)、Reuters (2013)、Checkoway et al. (2016)、Wikipedia "Dual_EC_DRBG"

スコープクリープと、非対称な適用

一度スキャン用のインフラが確立されれば、対象はCSAMに留まるとは限りません。「有害コンテンツ」の定義次第で、政治的反体制派が言及するURLも対象にできてしまいます。テロ対策、薬物、著作権――拡張の候補は常に列をなしています。

象徴的なのは、過去のChat Control案で、法執行機関・軍・情報機関・EUの一部省庁自身がスキャン義務の対象外とされていたことです。「ゴーイング・ダーク（捜査の暗転）」を懸念する側が、自らを監視対象から外そうとする――この非対称は、制度の本音を雄弁に語ります。

さらに、誤検知や差別的な誤り率、対象範囲の逸脱が起きたとき、誰が責任を負うのか。この「責任の所在」は、各国の法案がそろって沈黙している空白地帯です。

「グルーミング対策」論への反論 — 危険は主に身近にある

チャット検閲を正当化する最も強い論拠は、暗号化メッセージングを通じたグルーミング（性的搾取を目的とする手なずけ）への対策です。この危険は実在し、軽視すべきではありません。しかし児童性虐待の実態統計が一貫して示すのは、加害者の大多数が見知らぬネット上の他人ではなく、子どもが知っている人物だという事実です。米司法省の統計では、未成年被害者の9割以上が加害者と面識があり、3人に1人以上は家族・親族が加害者でした。

全市民のメッセージを盗聴しても、家庭という密室で起きる虐待は検知できません。むしろ誤検知の奔流が児童保護の現場からリソースを奪い（Part 3）、被害児童自身が相談や証拠の保全に使う安全な通信路を破壊します。虐待サバイバーの支援者の中からも、E2EEはむしろ被害者の生命線だという声が上がっています。

実効性のある児童保護とは、通報窓口・教育・児童相談体制・捜査リソースへの投資であって、全員の通信の盗聴ではありません。

出典：U.S. DOJ/BJS "Sexual Assault of Young Children as Reported to Law Enforcement"、WHO

焦点は9月へ — まだ何も決まっていない

Chat Control 1.0（任意スキャン）は2028年4月まで延命されましたが、E2EEサービスを対象から除外する修正が7月9日に採択されました。本丸のCSAR（Chat Control 2.0＝恒久規制）はトリローグが5回連続で決裂し、アイルランド議長国のもと、9月に第6ラウンドが予定されています。

最大の争点は、E2EEサービスにもクライアントサイド・スキャンを法的義務として課すか否かです。一部報道では、対象を「既知ハッシュとの照合」に限定し、AIによる未知コンテンツの分類は除外する妥協案も浮上していますが、この妥協が最終的に成立するかは依然として不透明です。そして本日見てきたとおり、「既知ハッシュとの照合」自体の信頼性が、いま根本から問われています。

出典：Andrea Fortuna（2026/7/10）、TechTimes、GIGAZINE、財経新聞（2026/7/11）

暗号は地政学である — 20法域の比較から

報告者が最近書いた論文 "The Geopolitics of Cryptography"では、E2EE規制をめぐる20法域の制度を比較し、規制のあり方を「禁止型・条件付きアクセス型・許容型・ハイブリッド型・ネット遮断依存型」の5類型に整理しました。

分析から見えたのは、各国の規制の動向を決めているのは必ずしも技術的制約ではない、ということです。違いを生むのは、「デジタル主権」をどう観念するか、政治体制の性格、そして合法傍受産業との制度的結びつきです。また暗号をめぐる規範形成では、プラットフォーム企業が国家に匹敵する「準主権的アクター」として振る舞い、その抵抗は市場規模と国家の強制力をにらんで選択的に行われています。グローバル・サウスへのネット規制の輸出も、こうした地政学的見地から捉える必要があります。

同論文はこのような理解を踏まえた規制設計の原則として、事前の技術的実現可能性評価、比例性、国際相互運用性、グローバルサウスを含むマルチステークホルダー統治の4つを導いています。Chat Controlの帰趨も、この地政学的文脈の中で読む必要があります。

出典：Hatta, "The Geopolitics of Cryptography" (ITSTokyo2026)

6

PART 6

代替案①：「手元で」完結する年齢認証

暗号を破壊せず、集めるデータを最小化する設計は可能です。すでに実装も動き始めています。

「年齢確認」ではなく「年齢の申立て」へ

Age Verification（現状）

訪問先のサイトやアプリごとに、IDや顔写真を直接収集する方式です。集める主体が増えるほど流出リスクが積み上がることは、Discordの事例が示す通りです。

Age Attestation（代替）

信頼できるプロバイダに一度だけ年齢を証明し、以降は「はい／いいえ」のトークンだけを各サイトに渡す方式です。本人が提供者を選び、自分の情報を管理できます。

暗号・プライバシー専門家のAlec Muffettは、この転換を「年齢確認（verification）」から「年齢の申立て（attestation）」への移行と整理します。適切な実装場所はOS・アプリストアの層であり、Meta CEOのMark Zuckerbergも2026年2月に「App Storeこそ適切な場所」と発言しました。情報をできるだけネットの「向こう側」に渡さず「手元」に留め、渡す情報を最小化する必要があります。

出典：Alec Muffett, Dropsafe（2026/3/4）

机上の議論ではない — 2026年、実装が動き始めている

Apple Declared Age Range API

生年月日を渡さず「年齢帯」だけをアプリに返します。2025年10月のベータを経て、2026年2月には豪州・シンガポール・ブラジル向けの18+アプリブロックへ拡張されました。

Google Play Age Signals API

同様の仕組みが2026年1月1日よりテキサス州で稼働を開始しました。ユタ州（5/6〜）とルイジアナ州（7/1〜）の州法が、実運用の試金石になります。

カリフォルニア州 AB1043（Digital Age Assurance Act）

年齢確認をOS・端末セットアップの時点にまで押し上げる計画です（2027年〜）。

英国 Ofcom

2027年1月までに、アプリストア単位での年齢確認についての報告書を公表する予定です。

出典：Apple Developer News、Techopedia（2026/5/20）、biometricupdate.com、median.co

Declared Age Range APIの仕組み ― 「宣言」であって「検証」ではない

この仕組みは、「ファミリー共有」内の「子どもアカウント（Child Account）」を前提とします。保護者が子どものApple IDをファミリーグループに接続すると、年齢帯の共有方針を「常に共有」「確認してから共有」「共有しない」の3つから選べます。アプリ開発者がDeclared Age Range APIを呼び出すと、返ってくるのは生年月日ではなく年齢帯（レンジ）と、その情報が「保護者による申告（guardianDeclared）」かどうかを示すフラグだけです。

重要なのは、この「年齢帯」がApple自身による本人確認の結果ではなく、子どもアカウント設定時に保護者が申告した情報にすぎないという点です。Appleは申告された年齢が正しいかどうかを検証していません。加えてAppleが渡すのは年齢という「シグナル」だけで、その年齢を各法域の年齢制限（例えば「13〜15歳」が何を意味するか）にどう翻訳し、何を制限するかは、個々のアプリ開発者の判断に委ねられています。

出典：Apple Developer（WWDC25 "Declared Age Range"）、Apple Newsroom（2025/6）

「手元」の代償 — 入口を支配する者が、参加そのものを支配する

年齢確認をOS・アプリストアの層に置く設計は、「アプリストア経由・OS完全性検証済みの経路でしかソフトウェアをインストールできない」という前提を暗黙のうちに伴います。報道によれば、EUの年齢認証アプリのAndroid版はGoogleの「Play Integrity API」への適合が条件となる見込みで、サイドローディングや代替ストア経由の配布はこの枠組みから排除されます。

この完全性検証の強化は、副産物として、「修理する権利」や、自分の端末に別のOSを入れる権利、サイドローディングの自由といった、EFFなどが長年擁護してきた権利と衝突します。EFFは修理する権利を重視し、「修理できないなら、それはもう自分のものではない」と警鐘を鳴らしてきました。年齢確認という大義名分が、この既存の攻防に新たな口実を与えています。

「手元」で完結させることは、集める情報の最小化ではあります。しかし実際に「手元」を支配しているのは利用者ではなく、OSベンダーとアプリストア運営者だという事実は変わりません。

出典：PhoneArena（2026）、Google Developer Blog、EFF "Defend Your Right to Repair"、Right to Tinker Initiative

業界内からの異論 — Signalの「Fusion」論

Signal CTOのEhren Kretは2026年3月、技術カンファレンス「Don't Be Evil」で各国の年齢確認規制を批判しました。「政策立案者が見落としているのは、本人確認（identification）と資格確認（verification）を切り離す義務を課していないことだ」。義務化するなら、プライバシー保護設計であることも同時に義務づけるべきだ、というのがKretの立場です。

Kretが根拠に挙げるのがゼロ知識証明などの匿名クレデンシャル技術です。「40年前からある論文を読めば済む話だ」と述べ、Signal自身の寄付システムを例に挙げます。「支払った人の集合に属する」ことだけを証明する領収書を発行し、誰が支払ったかは明かしません。年齢確認も同様に設計できる、というのが業界内で「Fusion」論と呼ばれる主張です。

OS・アプリストア発の「年齢シグナル」も同じ基準で検証すべきです。特定の端末アカウントに紐づいたまま渡される設計なら、それは匿名クレデンシャルでなく追跡可能なデータにすぎません。「手元で完結」と「プライバシー保護」は別の要件なのです。

出典：Ehren Kret (Signal CTO), FUTO "Don't Be Evil" Conference（2026/3）、Breitbart（2026/3/25）

銀の弾丸は無い

EUのデジタルIDウォレット（ゼロ知識証明ベース）も2026年、仏・デンマーク・ギリシャ・伊・西・キプロス・アイルランドの7カ国で試験導入されています。「18歳以上」であることだけを証明し、生年月日等を一切開示しない設計です。

それでも「ドアマット」論文の著者の一人であるSteven Bellovin（コロンビア大）は、プライバシー保護型のクレデンシャル方式にも「乗り越えがたい壁」が残ると指摘します。身分証そのものを取得できない層は、プライバシー保護型の仕組みからも取り残されます。発行機関の統治や運用コストの負担は未解決で、銀行口座開設や雇用確認など目的外利用への転用が始まれば、失効・取消の仕組みは権威主義的統制の道具にもなりえます。

アプリストア層への実装は、それ自体がプラットフォーム独占を強化するという懸念も浮上しています。年齢確認を「誰が」担うかを突き詰めれば、Apple・Googleという既存の門番の権限をむしろ拡大しかねません。だからこそ、規律の対象をもう一段深く、アーキテクチャそのものへ広げる発想が必要になります。

出典：Bellovin (2025)、digital-strategy.ec.europa.eu、biometricupdate.com

7

PART 7

代替案②：プロトコル・ベースのプラットフォーム規制

「何を見張るか」ではなく「誰が決めるか」を変える。監視に頼らずに情報空間の秩序をつくる、構造からのアプローチです。

規制の単位を変える ― 「プラットフォームではなくプロトコル」

ここまで見てきた規制は、年齢認証もチャット検閲も、「単一の中央集権的プラットフォーム」を所与とし、その振る舞いを外から矯正しようとするものでした。しかし数億件の投稿を単一の主体が裁定する中央集権的モデレーションは、「削除しすぎ」と「削除しなさすぎ」という相反する批判の板挟みから、原理的に抜け出せません。

Mike Masnickの「Protocols, Not Platforms」（2019）は、この前提自体を組み替える構想です。プロトコルとは〈誰もが従える、公開された共通の取り決め〉のこと。電子メールがSMTPという約束事の上でGmailからOutlookへ届くように、SNSも公開プロトコルの上に多数の競合する実装が乗る世界へ作り替える。一社がルールも入口も握る「囲い込まれた施設」から、開かれた「約束事」への転換です。

規律の対象を「コンテンツと個人」から「アーキテクチャ」へ移す。それがこの構想の核心です。

出典：Masnick, "Protocols, Not Platforms" (2019), Knight First Amendment Institute

フィルタの市場と、本人が鍵を握るデータストア

モデレーション権限は中央から末端へ

競合サービスや公益団体、地域コミュニティが提供する無数のフィルタをユーザ自身が選びます。問題ある言説は「完全に沈黙させず、リーチを絞る」形で扱われ、国家が「何を消すべきか」を裁定する必要は薄れます。

暗号化データストアの分離

投稿やソーシャルグラフは各ユーザが管理する「データバンク」に置かれ、サービス提供者はその都度アクセスを要求し、本人が範囲と期間を決めて許諾します。暗号化を破壊せず、本人同意を技術的な入口とするこの設計は、日本の「通信の秘密」とも整合的です。

唯一の大規模実装：Bluesky／AT Protocol

Blueskyでは、データ保管（PDS）・索引（Relay）・表示（App View）・ラベリング（Labeler）を層として分離し、ユーザはアカウントもフォロー関係も失わずに事業者を乗り換えられます。「言論の自由はあるが、リーチの自由はない」が運用の標語です。

もちろん課題もあります。フィルタ選択の複雑さ、需要側の取り込みの鈍さ、そして「分散プロトコルの中央集権的ガバナンス」という緊張は、Bluesky自身が現在進行形で示しているところです。

出典：Kleppmann et al. (2024)、The Bluesky Team (2024)、Hallinan et al. (2025)

年齢認証への応用 — 単一の「門番」に頼らない設計

二重非開示という設計 — euCONSENTの実証

EUの実証プロジェクトeuCONSENTのAgeAwareは、訪問先サイトが利用者の身元を知らず、年齢確認事業者もどのサイトを訪れたか知らない「二重非開示」を採ります。5カ国・約2,000人規模の実証で、相互運用の信頼性はほぼ100%でした。

一度の証明を、どこでも再利用する

SMTPが一社に縛られずGmailとOutlookをつなぐように、認定を受けた複数の年齢確認事業者が共通トークンを発行し、互いに検証し合います。利用者はサイトごとに何度もIDを出す必要がなく、事業者間の競争も保たれます。EUの公式「年齢認証アプリ」（p.9）が単一の中央集権的アプリであるのに対し、こちらは競合事業者が共通規格の上で相互運用するモデルです。

土台となるのはゼロ知識証明

Signal CTO Kretが求めた「本人確認と資格確認の分離」（p.45）は、この協調検証の上でこそ実装できます。生年月日そのものでなく「18歳以上である」という事実だけを、匿名クレデンシャルとして複数事業者の共通規格で標準化する——単一ベンダーの「手元」から、検証可能な相互運用へ。

Part 6で見た「手元」型の年齢認証は、データ収集を最小化する一方、実装の主導権をApple・Googleという既存の「門番」に委ねる代償を伴いました（p.44）。プロトコル化は、この主導権を単一ベンダーから相互運用する複数事業者へと組み替える、もう一つの解法です。

出典：Biometric Update（euCONSENT／AgeAware, 2022–2023）

CSAM・グルーミング対策への応用 — 層を分ける

インフラ層｜ホストできないものとして

AT Protocol（Bluesky）は、CSAMのような「ホストできない」コンテンツを、選べる任意のラベルでなく、インフラの不可避の義務として設計します。照合には業界横断の共有ハッシュDB（Thorn社Safer等）を用いる——PhotoDNA（Part3）と同じロジックです。

アプリ層｜選べる安全機能として

虐待の多くは家族・面識者によるもの（Part5, p.37）という実態を踏まえれば、必要なのは通信の一律スキャンではありません。利用者・保護者が選べる安全設定——不審接触の検知ラベル、未成年DMの制限、通報導線——を、暗号化を破らずアプリ層に用意することです。

チャット検閲（Part5）の核心的な誤りは、「公開・ホスティングの層」と「私的な通信の層」を混同し、両方に同じ手法——内容のスキャン——を適用しようとした点にあります。プロトコル的発想は、この二つの層を切り分けます。「何を見張るか」を国家が一律に決めるのではなく、「どの層で・誰が」対応するかを設計すべきです。

出典：Bluesky Moderation Architecture（2024）；Platformer（2024/11/26）

日本への実装 — 監視ではなく、構造を規律する

報告者は、日本の現行制度に接続する「三層併用」の立法案を提案しています。段階的に強制力を高める設計です。

入口（共同規制）

取引透明化法型 ｜ 経済産業省

指定大規模SNS事業者に相互運用性ロードマップの策定・公表を求め、国がモニタリング・評価します。

誘因（セーフハーバー）

情プラ法改正 ｜ 総務省

開かれたプロトコルを公開する事業者には、削除迅速化・透明化義務の一部履行とみなす「飴」を法定します。

後ろ盾（事前規制）

スマホ法拡張 ｜ 公正取引委員会

それでも動かない有力事業者には、指定と行為義務という事前規制を発動します。

年齢認証もチャット検閲も、規律の対象は「コンテンツと個人」でした。対象を「アーキテクチャ」へ移せば、国家のコンテンツ関与はむしろ減り、検閲との緊張も小さくなります。「すべてを見張る」ことなく秩序をつくる、もう一つの道です。

出典：八田「プロトコル・ベースのプラットフォーム規制—理論、海外動向、日本への実装」（近刊）

CONCLUSION

「AIに判断を委ねる」ことと「AIの判断に人間が責任を持ち続ける」ことは違う

実効性を検証せずに義務化し、誤りの責任を誰も引き受けない設計は、児童保護でもプライバシー保護でもありません。モラルパニックに立法を委ねるべきではありません。

暗号を破壊せず、集めるデータを最小化し、誤りに説明責任を持つ制度設計へ。年齢の証明は「手元で」完結させ、規律の対象はコンテンツからアーキテクチャへ移すべきです。

比例性・必要性・司法的説明責任――この3つの物差しが、これからの規制論議には必要です。

ご清聴ありがとうございました

質疑応答

八田 真行　駿河台大学 経済経営学部 ／ 日本IGF運営委員