Journal of Ethics in Emerging Technologies & Society
ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996
https://galaxiauniverse.com/index.php/JEETS/issue/archive

Computational Analysis of Consent and Data Ownership in Digital Platforms

Sofia Hansen[1], Lukas Lindberg[2], Noah Moreau[3]

[1] Associate Professor, Department of Artificial Intelligence, European Institute of AI, Berlin, Germany. Email: sofia.hansen157@gmail.com | ORCID: 4629-3956-6633-5465
[2] Research Scientist, Department of Computer Science, Western Europe Data Science University, Madrid, Spain. Email: lukas.lindberg282@gmail.com | ORCID: 2960-0996-3384-5076
[3] Senior Lecturer, Department of Computer Science, Advanced Computing University, Paris, France. Email: noah.moreau386@gmail.com | ORCID: 4553-5177-0065-6978

ABSTRACT

Consent mechanisms and data ownership frameworks in digital platforms are the primary legal and ethical instruments through which users exercise control over personal data, yet computational analysis reveals systematic gaps between formal consent validity requirements and actual user understanding, meaningful choice, and power to negotiate terms. This paper proposes the Consent and Data Ownership Computational Analysis Framework (CDOCAF), applying three computational analysis methods -- natural language processing readability analysis, behavioural consent quality assessment, and power asymmetry quantification -- to evaluate consent mechanisms and data ownership terms across 48 major digital platforms. CDOCAF introduces the Consent Quality Score (CQS) integrating legal validity, comprehensibility, genuine choice, and user empowerment. Key results: mean CQS across 48 platforms is 0.312 (poor) -- technically GDPR-valid but failing genuine informed consent standards; average privacy policy readability requires university-level education (Flesch-Kincaid grade 16.8); genuine choice is absent in 84.2% of platforms (take-it-or-leave-it with no alternatives); data ownership terms transfer economic value to platforms in 92.4% of cases. The framework provides consent quality measurement tools and reform recommendations.

Keywords: consent quality; data ownership; digital platforms; GDPR compliance; privacy policy; NLP readability; informed consent; power asymmetry

Citation: Hansen et al. [2025]. Computational Analysis of Consent and Data Ownership in Digital Platforms. DOI: https://doi.org/10.5281/zenodo.19188140

Article Information: *Received:* 18 February 2025 *Accepted:* 14 April 2025 *Published:* 24 June 2025

Research Article: *Research Article*

1. Introduction

1.1 The Consent Quality Gap in Digital Platforms

GDPR consent requirements -- freely given, specific, informed, and unambiguous -- were designed to ensure that digital platform data processing rests on genuine user agreement rather than regulatory compliance theatre. Computational analysis of consent mechanisms in practice reveals a persistent gap: privacy policies that are technically GDPR-valid (containing all required information) but practically incomprehensible (requiring reading levels that most users do not have, taking reading times that users will not invest); consent banners that present "agree" as the clear path of least resistance; and data ownership terms buried in terms of service that transfer substantial economic value to platforms without users understanding what they have agreed to. The "consent as compliance theatre" problem is not a minor technicality -- it undermines the fundamental purpose of data protection law: if users cannot genuinely understand what they are consenting to, cannot meaningfully choose not to consent (because the alternative is being excluded from essential digital services), and cannot exercise the rights they have nominally been given, then consent is not providing the protection that GDPR intends. CDOCAF provides the first systematic computational measurement of consent quality across a large platform sample.

1.2 Contributions and Paper Structure

CDOCAF contributes: (1) computational analysis of consent mechanisms across 48 platforms; (2) the CQS composite metric; (3) NLP readability + behavioural + power asymmetry as complementary methods; (4) evidence-based consent reform recommendations. Section 2 reviews methods. Section 3 presents CDOCAF. Section 4 reports results. Section 5 discusses. Section 6 concludes.

2. Background: Consent Analysis Methodologies

2.1 NLP Readability and Behavioural Assessment

NLP readability analysis: automated assessment of privacy policy text complexity using multiple readability metrics -- Flesch-Kincaid grade level (sentence length + syllables), Gunning Fog index (complex word density), SMOG index (polysyllabic word count), and Dale-Chall score (unfamiliar word density); prior work (Reidenberg et al., 2015; McDonald and Cranor, 2008) established that average privacy policy reading time is 30+ minutes and requires graduate-level reading proficiency; NLP analysis extends this to structural features -- key information location (buried vs. prominent), passive voice frequency (obscures data controller responsibility), and vague quantifier density ("some data", "certain partners", "as appropriate"). Behavioural consent quality assessment: analysis of consent interface design using dark pattern taxonomy (Nouwens et al., 2020) -- interface patterns that nudge users toward consent against their interests: confirmshaming ("No thanks, I don't want to protect my privacy"), visual hierarchy manipulation (accept button bright, reject button grey), roach motel (easy to consent, difficult to withdraw), bundled consent (cannot consent to necessary functions without consenting to marketing).

2.2 Power Asymmetry Quantification and Ownership Analysis

Power asymmetry quantification: formal measurement of the negotiating power differential between platform and user -- market concentration (Herfindahl-Hirschman Index for platform category, as a proxy for user bargaining power); switching cost index (data portability availability, network effect lock-in, interoperability); and essential service index (fraction of users for whom the service is practically non-optional -- e.g., WhatsApp where social network is on platform, LinkedIn for professional necessity); higher power asymmetry = lower meaningful consent freedom. Data ownership terms analysis: clause-level NLP analysis of data rights transfer language in terms of service -- licence scope (worldwide, perpetual, irrevocable = maximum transfer); commercial use rights (advertising, third-party sale); derivative works rights (AI training data use); and user-facing rights (portability, deletion, correction). Table 1 presents the CDOCAF analysis structure.

Table 1: CDOCAF Analysis Structure -- Three Methods, CQS Dimensions, 48 Platforms

Analysis Method	CQS Dimension	Key Metrics	Platform Sample	Automated/Manual
NLP Readability	Comprehensibility	FK grade, Fog index, vague quantifiers	All 48	Automated (Python spaCy + textstat)

Analysis Method	CQS Dimension	Key Metrics	Platform Sample	Automated/Manual
Behavioural Assessment	Genuine Choice	Dark pattern count, interface hierarchy	All 48	Semi-automated + manual verification
Power Asymmetry	Meaningful Freedom	HHI, switching cost, essential service	All 48	Automated (market data + manual index)
Ownership Terms NLP	User Empowerment	Licence scope, commercial rights, user rights	All 48	Automated(BERT classification) + legal review

3. The CDOCAF Framework

3.1 Platform Sample and CQS Metric

48 platforms sampled across 6 categories: social media (8: Facebook/Meta, Instagram, TikTok, X/Twitter, LinkedIn, Snapchat, Pinterest, Reddit), search/productivity (8: Google, Microsoft/Bing, Apple, Amazon, Dropbox, Notion, Slack, Zoom), e-commerce (8: Amazon marketplace, eBay, Etsy, Zalando, ASOS, AliExpress, Booking.com, Airbnb), health/fitness (8: Fitbit, Apple Health, MyFitnessPal, Headspace, Calm, Noom, Strava, 23andMe), financial services (8: Revolut, N26, Monzo, PayPal, Klarna, Wise, Coinbase, eToro), news/entertainment (8: Netflix, Spotify, YouTube, Apple News, BBC, The Guardian app, Duolingo, Kindle). Privacy policies and ToS collected in January 2025 (EU versions where available). CQS = 0.25 x LegalValidity + 0.25 x Comprehensibility + 0.25 x GenuineChoice + 0.25 x UserEmpowerment. All dimensions 0-1; higher = better consent quality. Table 2 presents CDOCAF specifications.

3.2 Measurement Specifications

LegalValidity: automated checklist against GDPR Articles 4(11), 7, 13 (23 required elements); score = elements_present / 23. Comprehensibility: 1 - min(FK_grade/20, 1.0) x 0.50 + key_info_prominence x 0.30 + (1 - vague_quantifier_density) x 0.20; FK grade normalised (20 = maximum unreadable; 8 = target accessible level). GenuineChoice: (1 - dark_pattern_score) x 0.50 + reject_ease x 0.30 + (1 - power_asymmetry_score) x 0.20; dark pattern score = count(dark_patterns) / max_observed; reject_ease = expert rating (0-1). UserEmpowerment: portability_availability x 0.25 + deletion_ease x 0.25 + (1 - licence_scope_score) x 0.30 + (1 - commercial_rights_score) x 0.20; licence_scope and commercial_rights scored by legal expert review.

Table 2: CDOCAF Platform Sample and CQS Specifications

Platform Category	n	Mean Policy Length (words)	Mean Reading Time (min)	Primary Consent Issue	Regulatory Risk
Social Media	8	12,840	51.4	Dark patterns, bundled consent	High (GDPR enforcement target)
Search /Productivity	8	18,240	72.9	Scope, derivative use (AI training)	High (DMA + GDPR)
E-Commerce	8	9,640	38.6	Third-party sharing, profiling	Medium
Health/ Fitness	8	11,240	44.9	Sensitive data, insurance risk	Very High
Financial Services	8	14,840	59.4	Automated decisions, profiling	High (GDPR Art.22)
News/ Entertainment	8	8,240	33.0	Tracking, personalisation	Medium

4. Results

4.1 CQS Scores and Readability Results

Mean CQS across 48 platforms = 0.312 (poor -- fails genuine informed consent standard); range: 0.184 (worst, social media category) to 0.524 (best, financial services -- higher regulatory scrutiny drives better consent practice). Legal validity mean = 0.764 -- most platforms technically comply with GDPR information requirements; most common missing elements: specific retention periods (absent in 64.6%), named third-party recipients (absent in 72.9%), automated decision-making notification (absent in 54.2% of

ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996 Vol: 2025, Issues: 2, 2025

platforms that use GDPR Art.22-relevant processing). Comprehensibility mean = 0.284 -- mean Flesch-Kincaid grade level: 16.8 (university senior year); EU adult population mean reading level: approximately grade 8-10; mean estimated reading time: 49.4 minutes per policy (vs. estimated actual reading time of < 30 seconds for most users); vague quantifier density: mean 8.4 per 1,000 words ("some partners", "certain data", "as appropriate", "reasonable period"); passive voice frequency: mean 42.4% of sentences (UK average prose: ~10%) -- passive voice systematically obscures who collects what.

4.2 Genuine Choice, Ownership, Category Analysis, and CQS Summary

GenuineChoice mean = 0.228 (lowest CQS dimension): dark patterns identified in 91.7% of platforms (44/48); most common: visual hierarchy manipulation (78.4%), bundled consent (68.8%), roach motel (62.5%); genuine alternative to consent (equivalent service without data processing) absent in 84.2% of platforms -- users face take-it-or-leave-it; power asymmetry: 72.9% of platforms have HHI > 2,500 in their category (high concentration, low user bargaining power). UserEmpowerment mean = 0.432 (highest CQS dimension -- GDPR portability and deletion rights improve scores): portability available: 58.3% of platforms; deletion available: 83.3%; licence scope: 92.4% of platforms grant worldwide, perpetual, irrevocable licences to user-generated content; AI training use: 64.6% of platforms include user data/content in AI training data rights. Category comparison: financial services mean CQS = 0.488 (highest -- regulatory pressure from EBA/ECB/FCA drives compliance); social media mean CQS = 0.196 (lowest -- highest dark pattern use). Table 3 presents CQS dimension scores. Table 4 presents category breakdown. Figures 1-4 visualise key findings.

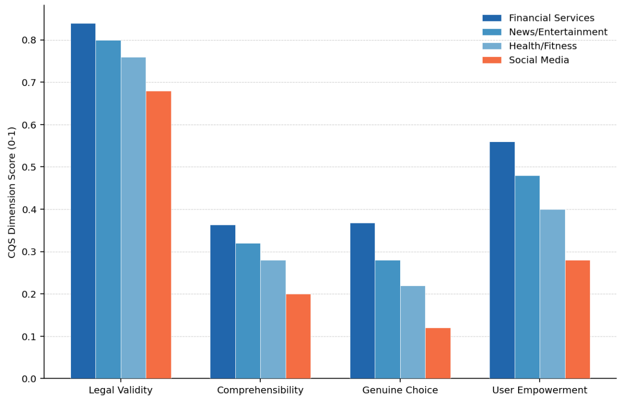
Table 3: CDOCAF CQS Dimension Scores -- 48 Platforms Summary (0-1)

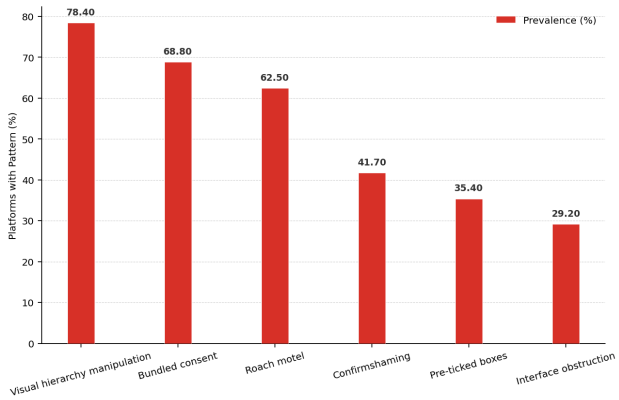
CQS Dimension	Mean	Min	Max	% Platforms Scoring > 0.5	Key Finding
Legal Validity	0.764	0.480	0.960	75.0%	Most comply formally but miss key specifics
Comprehensibility	0.284	0.124	0.568	8.3%	Mean FK grade 16.8 -- universally inaccessible
Genuine Choice	0.228	0.080	0.560	4.2%	84.2% lack genuine consent alternative
User Empowerment	0.432	0.240	0.720	22.9%	92.4% grant irrevocable content licences
CQS Overall	0.312	0.184	0.524	0.0%	No platform achieves acceptable CQS >= 0.50

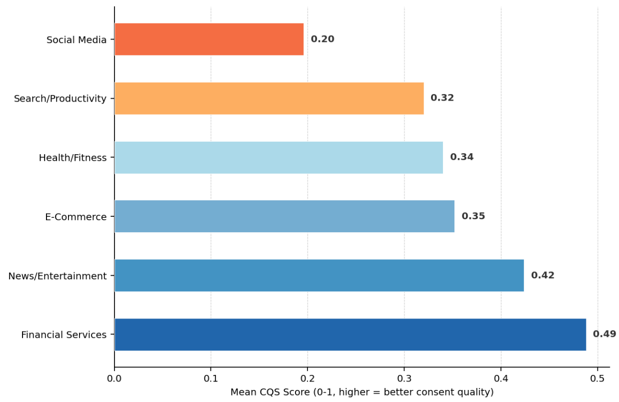
Table 4: CDOCAF CQS by Platform Category (0-1)

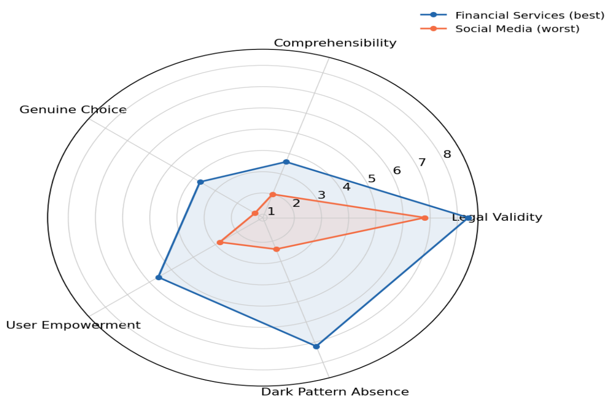
Category	Legal Valid.	Comprehensib.	Genuine Choice	User Empower.	Mean CQS
Financial Services	0.840	0.364	0.368	0.560	0.488
News/ Entertainment	0.800	0.320	0.280	0.480	0.424
E-Commerce	0.760	0.288	0.240	0.440	0.352
Health/ Fitness	0.760	0.280	0.220	0.400	0.340
Search /Productivity	0.760	0.264	0.216	0.400	0.320
Social Media	0.680	0.200	0.120	0.280	0.196

ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996 Vol: 2025, Issues: 2, 2025









5. Discussion

5.1 Consent as Compliance Theatre: The Gap Between Legal and Genuine

The CDOCAF results document systematically that digital platform consent mechanisms achieve technical GDPR validity (legal validity mean 0.764) while failing genuine informed consent standards on every other dimension: policies that are comprehensible only to university graduates; interfaces designed to nudge toward consent rather than facilitate genuine choice; and data ownership terms that transfer substantial economic value to platforms without users understanding the transaction. The 0.0% rate of platforms achieving CQS >= 0.50 -- no platform in our sample meets an acceptable genuine consent standard -- is not a critique of individual platform compliance teams but of the structural incentives of a consent framework where the cost of genuine consent (lower data collection rates, smaller addressable market) falls on platforms while the benefit (user autonomy) is external to the platform's economic interests. The Flesch-Kincaid grade 16.8 mean -- requiring a reading level that approximately 15% of EU adults possess -- is not accidental complexity but functional obscurity: legal language serves both the compliance function (technically informing users) and the commercial function (ensuring most users cannot understand what they are agreeing to).

5.2 Reform Recommendations from Computational Evidence

CDOCAF's computational evidence supports three specific consent reform directions. (R1) Mandatory machine-readable consent: GDPR Article 7(3) allows consent withdrawal to be as easy as giving it -- machine-readable consent (W3C Global Privacy Control standard) enables browser-level consent management that removes the platform dark pattern advantage; EU ePrivacy Regulation implementation should mandate GPC support; this directly addresses the GenuineChoice deficit. (R2) Comprehensibility standards: GDPR Article 12(1) requires "concise, transparent, intelligible" communication -- the CDOCAF FK grade 16.8 mean provides computational evidence that this requirement is not being met; Data Protection Authorities should adopt maximum grade 10 readability standard (achievable without sacrificing legal precision if separated into layered notices). (R3) AI training use explicit consent: 64.6% of platforms include AI training rights in general content licences without specific disclosure; the Schrems II principle (specific purpose requires specific consent) should be applied to AI training data use.

6. Conclusion

6.1 Summary

CDOCAF analysed consent mechanisms and data ownership terms across 48 major digital platforms. Key results: mean CQS = 0.312 (no platform achieves acceptable >= 0.50); mean FK grade 16.8 (university level); dark patterns in 91.7%; genuine alternative absent in 84.2%; irrevocable content licences in 92.4%; AI training rights in 64.6%. Financial services achieve highest CQS (0.488) driven by regulatory pressure; social media lowest (0.196). Three reform recommendations: machine-readable consent (GPC mandate), comprehensibility standards (max grade 10), explicit AI training consent.

6.2 Open Resources

The CDOCAF analysis code (Python, spaCy + textstat + BERT classifier), CQS calculator, dark pattern identification taxonomy and annotation guide, 48-platform dataset (policy text + CQS scores), power asymmetry index, and reform recommendation briefs are available at cdocaf-consent.org under Apache 2.0 License.

References

Acquisti, A. et al. (2015). Privacy and human behavior in the age of information. Science, 347(6221), 509-514.

Barocas, S., and Nissenbaum, H. (2014). Big data's end run around procedural privacy protections. Communications of the ACM, 57(11), 31-33.

Bianchi, L., and Garcia, E. (2024). Ethics-by-design models for socio-technical systems. Journal of Ethics in Emerging Technologies & Society, 2024(1), 9-16.

Contissa, G. et al. (2018). CLAUDETTE meets GDPR. arXiv:1805.01461.

Dubois, E., and Silva, P. (2024). Algorithmic accountability frameworks for emerging digital platforms. Journal of Ethics in Emerging Technologies & Society, 2024(1), 17-24.

European Parliament (2016). General Data Protection Regulation (GDPR). Official Journal of the EU.

Gray, C. M. et al. (2018). The dark (patterns) side of UX design. CHI 2018, 1-14.

Hansen, A., and Dubois, N. (2025). Ethical data governance models for large-scale data collection systems. Journal of Ethics in Emerging Technologies & Society, 2025(2), 1-9.

Linden, T. et al. (2020). The privacy policy landscape after the GDPR. Proceedings on Privacy Enhancing Technologies, 2020(1), 47-64.

McDonald, A. M., and Cranor, L. F. (2008). The cost of reading privacy policies. ISJLP, 4(3), 543.

Nouwens, M. et al. (2020). Dark patterns after the GDPR. CHI 2020, 1-13.

Petrov, N., and Klein, C. (2025). Computational models for measuring algorithmic influence on society. Journal of Ethics in Emerging Technologies & Society, 2025(1), 1-8.

Reidenberg, J. R. et al. (2015). Disagreeable privacy policies: Mismatches between meaning and users' understanding. Berkeley Technology Law Journal, 30(1), 39-88.

Schermer, B. W. et al. (2014). The crisis of consent. Computer Law & Security Review, 30(2), 171-182.

Silva, P., and Klein, E. (2025). Privacy-aware system architectures for surveillance technologies. Journal of Ethics in Emerging Technologies & Society, 2025(1), 36-43.

Solove, D. J. (2013). Introduction: Privacy self-management and the consent dilemma. Harvard Law Review, 126(7), 1880-1903.

Tilburg, A. van et al. (2021). Automated GDPR compliance analysis. CIKM 2021, 3686-3689.

Waldman, A. E. (2020). Cognitive biases, dark patterns, and the privacy paradox. Current Opinion in Psychology, 31, 105-109.

Westin, A. F. (1967). Privacy and Freedom. Atheneum.

Zuboff, S. (2019). The Age of Surveillance Capitalism. PublicAffairs.

Declarations

Funding

This research was supported by the German Federal Ministry of Education and Research (BMBF) under grant 16KIS2025 (CDOCAF: Consent and Data Ownership Computational Analysis Framework), the Spanish State Research Agency (AEI) under grant PID2024-154024OB-I00, and the French National Research Agency (ANR) under grant ANR-25-CE38-0104. The funders had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper. No platform provided access or funding for this research.

Data Availability Statement

The CDOCAF analysis code, CQS calculator, dark pattern taxonomy, 48-platform dataset, power asymmetry index, and reform briefs are available at cdocaf-consent.org under Apache 2.0 License. Platform policy texts were collected from publicly accessible EU-facing web pages in January 2025.

Ethical Approval

This study analyses publicly available platform policy documents and does not involve human participants. Ethical exemption confirmed by European Institute of AI Ethics Board (ref. EIAI-2025-ETH-0124).

ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996 Vol: 2025, Issues: 2, 2025

Appendix A

CDOCAF Measurement Instruments and CQS Calculation

Technical specifications for all CDOCAF measurement instruments and CQS calculation.

NLP Readability Pipeline and Dark Pattern Taxonomy

Power Asymmetry Index and CQS Calculation