EU/EC GDPR Audit Mechanism: VDaaS
Track 7: Technology Innovation

Vinden Wylde*, Edmond Prakash, Chaminda Hewage, and Jon Platts

Keywords: GDPR, Bias, Auditing, Digital Services Act, Big Data, AI

1 Introduction

Of the lessons learned from Covid-19, the unprecedented global public health emergency coupled with mandatory governmental requirement for the public to stay at home, put significant users' information bias and data privacy at a heightened risk of violation as a consequence. In [1] the authors consider integrity violations of social media posts, individuals, groups, and advertisers of social networks, that potentially were used as a vehicle in violating policies such as present and future exploitation of children (i.e., grooming). As messaging services head toward end-to-end encryption, the balance between privacy and encryption of private posts are briefly discussed.

However, the General Data Protection Regulations 2016/679 (GDPR) article 7 and recital 32, requires that consent must be freely given, specific, informed, and unambiguous. This means that with any cookie banner for example, that if there is any false or misleading information then this will violate GDPR.

Vinden Wylde
Cardiff School of Technologies, Cardiff Metropolitan University, CF5 2YB
e-mail: vwylde@cardiffmet.ac.uk

Edmond Prakash
Cardiff School of Technologies, Cardiff Metropolitan University, CF5 2YB
e-mail: eprakash@cardiffmet.ac.uk

Chaminda Hewage
Cardiff School of Technologies, Cardiff Metropolitan University, CF5 2YB
e-mail: chewage@cardiffmet.ac.uk

Jon Platts
Cardiff School of Technologies, Cardiff Metropolitan University, CF5 2YB
e-mail: jplatts@cardiffmet.ac.uk

An analysis of cookie banners that included the assignment of the correctness of cookie-to-category shown, the claimed cookie expiration time, and the overall completeness of the cookie banner, was undertaken by [2] to inspect the accuracy of said declarations. Their results showed that many website ambiguities occurred to include relabelling the same cookie many times for different and/or contradictory purposes, and with additional undeclared, sometimes unclassified cookies.

In addition, an article by [3] pertains to the Digital Services Act (DSA) for Big Tech platforms (i.e., Instagram and Facebook [Meta], and Youtube [Google]) in the assessment and management of systemic risk regarding their services. This includes risks such as the spread of misinformation and advocacy of hatred, and means that in a "watershed moment" for internet regulation, that Big Tech companies will have to present annual independently verified audits [4], give platform access to civil society, regulator and third-party researchers [5], and present insights into algorithmic "black-box" accountability, thus enabling greater scrutiny.

Therefore, this study proposes a hybrid intelligent data enforcement pipeline user interface (UI) to aide in the identification and detection of data content violation/bias in big data (as opposed social media content), and aims to utilise GDPR and bias (i.e., privacy vs ethics) frameworks in assessing the efficacy of producing an auditing and validation mechanism (Fig. 1).

This task will require the identification of specific metrics to enable the tracking and ultimate assessment, to inform users, shareholders, and regulators, in the designing and implementation of transparent and accountable applications to relevant populations. The following challenges concern the recent EU/EC DSA, and the potential violations of messaging services and transparency of websites.

2 Research Challenges

2.1 Free Expression and Safety: Social Media Posts

In [1] the authors consider the occurrences of integrity violation regarding social media posts to include those of paid advertisements, groups, and individuals. With classification, semantics (text) developments and advances include supervised training (see [6], [7], and [8]), with standard text understanding architectures such as [6] (see also [9], [10], and [11]), detecting emotion [12], potential lies [13], and fake news [14]. Additional challenges are identified in how the integrity of published community policies (i.e., companies and the European Commission) are defined and applied on social media platforms in meeting the "right balance" in keeping the platform safe and to facilitate free expression.

2.2 Data Collection Consent: Cookies

In [2], the authors highlight the European Union's GDPR, and the requirement for websites to request consent and to inform users of personal data collection through cookies. However, in most websites there are no choices offered, whilst other websites attempt to deceive users into accepting all cookies. The authors highlight and document the severity of this situation, by analysing potential GDPR violations from around 30k website cookie banners. Six novel categories of violations are identified to include misleading expiration times and incorrect category assignments (see also [15], [16], and [17]), thus identifying a minimum of one violation in 97.7% of the analysed websites sampled.

3 Research Actions and Possible Solutions

3.1 Methods of Integrity Enforcement: Main Challenges

In assessing technique efficacy for the identification of data/content violation, a set of tracking metrics are needed. However, in contrast to the adversarial nature of integrity (unlike machine learning applications), and the low occurrences of particular violations, this can present additional challenges in the designing of appropriate metrics. However, as a metric, prevalence for example, can be utilised (i.e., document recall of web documents), if not somewhat difficult, therefore in using prevalence would mean to actually count network posts/data that are distinct. In addition, due to posts and the frequencies of revisiting by a user, it may be more beneficial to indicate and measure of "bad experiences" of potential violating posts, and to include experience prevalence as an additional measure to indicate severity [1].

Big Data: Data Controller Emulation. Here, the identifying of data veracity types (data cleansing) is important to defining data type metrics for recording and processing purposes. For example, additional work is necessary, particularly in the areas of granting a users the right to edit, update, and delete their data (Article 16). Also, there is also a need for policies designed to give clear definitions of data storage methods, to obtain a comprehensive and integrated view of what personal data is using for storage from within an organisation context, and to ensure the production of the necessary records from data processing activities.

Ethics: Categorisation of Principles, Rights, and Freedoms. Next, we undertake (Data Protection Officer) an analysis of prevalence in the context of upholding Integrity and Confidentiality principles; processing being Lawful, Fair, and Transparent (Article 5(1)), Storage Limitation (Article 5(1)(b)), and rights concerning racial or ethnic origin, political opinions, religion (Recital 75).

Data Protection: Categorisation of Bias and Violations. Finally, a further analysis of legal framework (i.e., GDPR) with a view to applying technological mechanisms.

The relevant policies regarding the data subject will be transcribed, alongside allocating appropriate metrics will be applied for data collection. Information such as: Measures taken in addressing breach, describe nature of personal data breach, and to describe possible effect of breach for individual will be quantified.

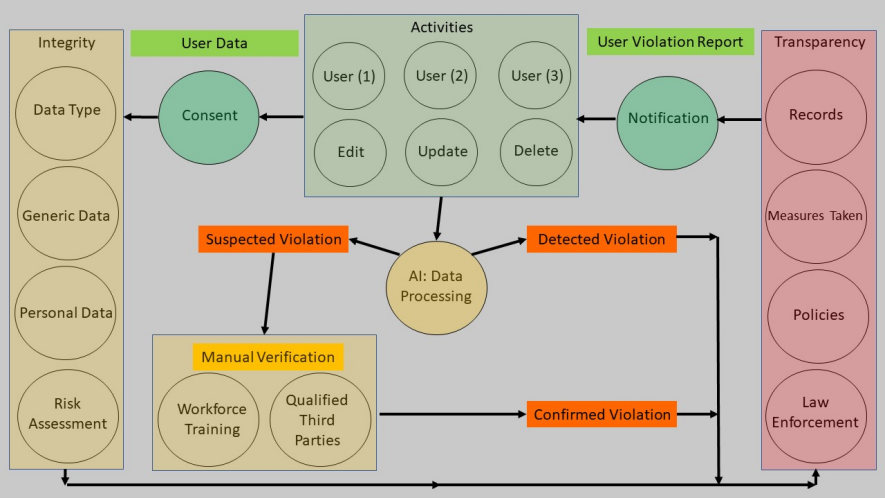

Fig. 1 VDaaS Audit Framework

4 Conclusion

This research looks at a sample of management systems and emerging technologies to ascertain challenges linked to include data privacy, risk severity, user consent, and information bias. In the development stages of this project, the findings will provide leverage and contribution to further research, organisation change, and cultural policy development. This will also further enable the protections of GDPR, augmented by AI, thus upholding its application to user rights and freedoms in respect to overall internet governance.

References

[1] A. Halevy, C. Canton-Ferrer, H. Ma, U. Ozertem, P. Pantel, M. Saeidi, F. Silvestri, and V. Stoyanov. Preserving Integrity in Online Social Networks. *Communications of the ACM*, 65(2):92–98, 2022.

[2] D. Bollinger, K. Kubicek, C. Cotrini, and D. Basin. Automating Cookie Consent and GDPR Violation Detection. In *31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, 2022.

[3] European Union: Digital Services Act Agreement a 'watershed moment' for Internet Regulation. `https://www.amnesty.org/en/latest/news/2022/04/european-union-digital-services-act-agreement-a-watershed-moment-for-internet-regulation/`. Accessed 01.05.2022.

[4] A. Peukert, M. Husovec, M. Kretschmer, P. Mezei, and J. Quintais. European Copyright Society-Comment on Copyright and the Digital Services Act Proposal. *Available at SSRN 4016208*, 2022.

[5] A. Turillazzi, F. Casolari, M. Taddeo, and L. Floridi. The Digital Services Act: An Analysis of its Ethical, Legal, and Social Implications. *Legal, and Social Implications (January 12, 2022)*, 2022.

[6] J. Devlin, M. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[7] M.E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. Deep Contextualized Word Representations. arXiv preprint arXiv: 180205365, 2018.

[8] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R.R. Salakhutdinov, and Q.V. Le. Xlnet: Generalized Autoregressive Pretraining for Language Understanding. *Advances in neural information processing systems*, 32, 2019.

[9] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. Roberta: A Robustly Optimized Bert Pretraining Approach. *arXiv preprint arXiv:1907.11692*, 2019.

[10] Z. Lan, M. Chen, S. Goodman, K. Gimpel, and R. Sharma, P.and Soricut. Albert: A Lite Bert for Self-Supervised Learning of Language Representations. *arXiv preprint arXiv:1909.11942*, 2019.

[11] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P.J. Liu. Exploring the Limits of Transfer Learning With A Unified Text-to-Text Transformer. *arXiv preprint arXiv:1910.10683*, 2019.

[12] S. Rajamanickam, P. Mishra, H. Yannakoudakis, and E. Shutova. Joint Modelling of Emotion and Abusive Language Detection. *arXiv preprint arXiv:2005.14028*, 2020.

[13] R. Mihalcea and C. Strapparava. The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language. In *Proceedings of the ACL-IJCNLP 2009 conference short papers*, pages 309–312, 2009.

[14] C.L.M. Jeronimo, L.B. Marinho, C.E.C. Campelo, A. Veloso, and A.S. da Costa Melo. Fake News Classification Based on Subjective Language. In

Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services, pages 15–24, 2019.

[15] X. Hu, N. Sastry, and M. Mondal. CCCC: Corralling Cookies into Categories with CookieMonster. In *13th ACM Web Science Conference 2021*, pages 234–242, 2021.

[16] X. Hu, G.S. de Tangil, and N. Sastry. Multi-Country Study of Third Party Trackers From Real Browser Histories. In *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 70–86. IEEE, 2020.

[17] C. Matte, N. Bielova, and C. Santos. Do Cookie Banners Respect My Choice?: Measuring Legal Compliance of Banners From IAB Europe's Transparency and Consent Framework. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 791–809. IEEE, 2020.