Out of Unmündigkeit: The GDPR for User Enlightenment

[1]**Malin Eiband,** [1]**Hanna Schneider,** [2]**Fabius Steinberger**
[1]LMU Munich, Munich, Germany
[1]{malin.eiband, hanna.schneider}@ifi.lmu.de
[2]fabius.steinberger@acm.org

Abstract

Data processing and machine learning have advanced technology in the Information Age, yet users' *digital literacy* remains in its infancy. We argue that this discrepancy is due to a lack of transparency in current systems, which keeps users in a state of Unmündigkeit (immaturity) and thus unable to make informed choices when using personalized products and services. We see EU's General Data Protection Regulation as the legal framework and a unique opportunity to foster such literacy and user enlightenment. In the light of our call, we present possible research avenues for the HCI community to leverage this opportunity.

Author Keywords

GDPR; Digital Literacy; Transparency; Interpretability; Accountability; Scrutability; System Design.

ACM Classification Keywords

H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

Introduction

The Information Age epitomizes an economy based on the exchange of data. In particular, *personal data* have been described as the new oil or the currency of the twenty-first century [8]. Data mining and machine learning transform this precious resource into new, personalized services

and products. Yet, despite considerable research efforts (e.g., [1, 6, 20, 28, 33]), privacy concerns remain an ongoing challenge. For example, it has been reported that personal data processing evokes indistinct feelings of "creepiness" or helplessness [27]. Nevertheless, studies have demonstrated that users do provide their personal data in exchange for (even small) benefits [4]. This discrepancy between attitudes and behavior is known as the privacy paradox [24]. In this position paper, we draw on work by Hildebrandt [11] and argue that it may be caused by *insufficient literacy* in terms of data processing. This digital literacy has not kept pace with quick technological advances. As a result, users are often unable to understand, assess, and evaluate the impact of algorithmic decision-making. They therefore depend on the guidance and benevolence of service providers, wiping away concerns when giving in to the promised benefits, simplicity, and convenience in exchange for their data.

Unmündigkeit, Digital Literacy and Transparency

To us, this situation is reminiscent of what Kant describes in his response to *"What is Enlightenment?"* [12]. Kant sharply criticizes the "laziness and cowardice" of men in terms of using their own understanding. However, in contrast to Kant who above all holds men themselves responsible for their Unmündigkeit (immaturity), we find evidence that this time such Unmündigkeit might *not* be self-imposed: It has been shown that users are very much interested in the underlying reasoning and processing of systems [9, 18], and that users are able to establish accurate mental models of their workings [30].

Instead, we argue that it is the dependency of users on the goodwill of providers that hinders user enlightenment. In his essay, Kant describes the relationship between men and their "guardians", i.e., representatives of social institutions, "who have kindly taken supervision upon themselves" and thus undermine men's ability to step out of Unmündigkeit. Similarly, users currently lack the *opportunity* to leave the state of Unmündigkeit because most systems remain opaque and providers are reluctant to make them transparent. This reluctance may be due to additional efforts or costs required for more transparency. It may also be due to business secrecy, or because it is yet often unclear what "making a system transparent" means in practice. Or, in the context of neural networks and deep learning, because the feasibility and ways of making intelligent algorithmic processing transparent are still to be established (e.g., [26]).

A lack of transparency and understanding of personal data processing can discourage users from utilizing and enjoying digital services and products. For example, Felt et al. [10] have shown that users tend to refrain from installing apps because of privacy concerns. Opaqueness may also lead to decreased user understanding due to erroneous mental models of the system [16, 30].

The General Data Protection Regulation (GDPR) is the new legal framework aimed at protecting data privacy within the European Union. A major requirement for GDPR compliance is transparency. We therefore argue that we are now presented a unique opportunity for user enlightenment. We can leverage this opportunity to foster digital literacy by making data processing and algorithmic decision-making transparent to users. This is in line with prior calls for more transparency in data processing and algorithmic decision-making [13, 32]. However, this time, the legal prerequisites are different.

Prior work has associated transparency and information about algorithmic decision-making with increased user trust in the system [7] and satisfaction with its recommenda-

tions [5]. Moreover, transparent systems help users build better mental models of their workings [15]. We therefore encourage companies and providers to not only see the GDPR and transparency as a burden, but as a strategic opportunity of enhancing products and services and thus increasing customer loyalty. With this call we position ourselves among researchers such as Mulligan and Bamberger who made a related argument in the context of privacy-by-design [23].

We furthermore argue that certificates, labels or seals demonstrating GDPR compliance alone might not be sufficient – they might establish user trust, but will do little to advancing digital literacy and user enlightenment. In contrast, we call for profound yet usable information as a means to fulfill the requirements regarding transparency in the GDPR and to give users the opportunity to step out of their current Unmündigkeit through informed choices – as Kant puts it:

> "[...] that the public should enlighten itself is more likely; indeed, if it is only allowed freedom, enlightenment is almost inevitable." [12]

To reach this goal, we see the HCI community as a mediator between industry and users to finding answers to the question:

> How can we anchor enlightenment in design?

In the following sections, we will draw on related work to identify possible avenues for future research towards best practices for user enlightenment.

Towards Enlightenment in Design

There is a plethora of work on transparency in various areas such as interactive [2] and interpretable machine learning [31], recommenders [25], or context-aware systems [18]. We review prior work in the light of our call for user enlightenment and group it according to the two main challenges that we see for putting this call into design.

First Challenge: What Kind of Information Allows for Informed Choice?

Making informed choices about one's privacy implies being provided with (exactly) the pieces of information that allow one to do so. What sounds trivial at first sight is a complex task when it comes to making algorithmic decision-making transparent. A common approach in prior work are so-called *explanation interfaces*, which typically answer possible questions users might have about the workings and decisions of the system. Lim and Dey [18], for example, introduced five types of questions in context-aware systems (e.g., *Why* did the system do X? *Why* did it *not* do Y? *What (else)* is it doing?), which have been adopted by, for example, Kulesza et al. [15]. In the same manner, Miller [21] defines explanations as answers to *Why* questions. Drawing on humanities, he provides a very detailed overview of the prerequisites for a "good" explanation (as summarized by [22]):

(1) *Contrastiveness*: Explain why one prediction was made *instead of another* prediction.
(2) *Selectiveness*: Select information instead of explaining *all* possible causes for a prediction.
(3) *Sociality*: Converse with the receiver of the explanation.
(4) *Abnormality*: Focus on abnormal or unusual causes.
(5) *Truthfulness*: Create explanations that are true and hold true also in other situations.

(6) *Coherence*: Explain in line with prior beliefs of the receiver of an explanation.

(7) *Generalizability and probability*: In the absence of an abnormal cause, explain in a general and transferable manner.

However, it is yet unclear how to determine the *actual* content of an explanation. *How can we make sure that the explanation we give is suited to foster digital literacy and thus allows users to make informed choices? What is the ideal granularity of an explanation?* In line with [22], we argue that this is dependent on the application context and user group alike, and thus has to be defined for each system or system group separately. We encourage the HCI community to explore ways of finding answers to these questions to provide companies and providers with approaches they can apply without too much effort. Otherwise, we see a risk of one-fits-all solutions for transparency, similarly to what we have seen in end user licence agreements: They are straight-forward for companies, but do not take into account the needs of users [3]. For example, following the approach by Kulesza et al. [16, 17] and Tullio et al. [30], Eiband et al. [9] have proposed a process driven by users' mental models to make system decisions transparent. However, other approaches may exist.

Second Challenge: Which Presentation Format Facilitates Informed Choice?
We argue that the effectiveness of explanations in terms of user enlightenment is strongly dependent on the way it is presented; presentation should allow users to efficiently grasp and process the given information in a particular context. *But which presentation formats are suited for efficient processing and how can we find them in a structured way?* Although prior work has already explored different ways of presenting information, there are yet no agreed-upon approaches to this problem, nor best practices. One notable

exception is a toolkit for automatically creating standardized text-based explanations in context-aware systems introduced by Lim and Dey [19]. Beyond that, researchers have tested text-based explanations as well as multimedia visualizations for various systems. For example, while Lim and Dey recommend visualizations to augment explanations in context-aware systems [18], Kouki et al. found that text-based explanations performed similarly well as Venn diagrams in recommenders [14]. We argue that digital literacy and informed choice might be best fostered if the presentation format is standardized to a certain extent across products and services, similar to existing design patterns for graphical user interfaces [29]. These patterns should be designed in a way that takes company-specific requirements into account while creating a "vocabulary" of interface elements that users can quickly understand and recognize in similar situations.

Concluding Remarks

In this position paper, we have argued that the GDPR presents a unique opportunity for the HCI community to foster users' digital literacy through making data processing and algorithmic decision-making transparent. To us, the lack of this data literacy results in users' dependency on the goodwill of providers, which is reminiscent of the relationship between men and their "guardians" criticized by Kant [12]. We therefore see digital literacy as a *necessary* step for users to step out of their current Unmündigkeit, and call upon HCI researchers to find design solutions to pave the way for user enlightenment. Whether it is indeed *sufficient* for users to become enlightened will become clear if we give them the opportunity to do so.

REFERENCES

1. Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck,

Lorrie Faith Cranor, and Yuvraj Agarwal. 2015. Your Location Has Been Shared 5,398 Times!: A Field Study on Mobile App Privacy Nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 787–796. DOI: `http://dx.doi.org/10.1145/2702123.2702210`

2. Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine* 35, 4 (2014), 105–120.

3. Omri Ben-Shahar. 2009. The Myth of the "Opportunity to Read" in Contract Law. *European Review of Contract Law* 5, 1 (2009), 1–28.

4. Juan Pablo Carrascal, Christopher Riederer, Vijay Erramilli, Mauro Cherubini, and Rodrigo de Oliveira. 2013. Your Browsing Behavior for a Big Mac: Economics of Personal Information Online. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13)*. ACM, New York, NY, USA, 189–200. DOI: `http://dx.doi.org/10.1145/2488388.2488406`

5. Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. 2008. The Effects of Transparency on Trust in and Acceptance of a Content-based Art Recommender. *User Modeling and User-Adapted Interaction* 18, 5 (20 Aug 2008), 455. DOI:`http://dx.doi.org/10.1007/s11257-008-9051-3`

6. Justin Cranshaw, Jonathan Mugan, and Norman Sadeh. 2011. User-controllable Learning of Location Privacy Policies with Gaussian Mixture Models. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI'11)*. AAAI Press,

1146–1152. `http://dl.acm.org/citation.cfm?id=2900423.2900605`

7. Mary T. Dzindolet, Scott A. Peterson, Regina A. Pomranky, Linda G. Pierce, and Hall P. Beck. 2003. The Role of Trust in Automation Reliance. *International Journal of Human-Computer Studies* 58, 6 (2003), 697 – 718. DOI:`http://dx.doi.org/https://doi.org/10.1016/S1071-5819(03)00038-7`

8. The Economist. 2017. Regulating the Internet Giants: The World's Most Valuable Resource is no Longer Oil, but Data. (May 2017). Retrieved February 7, 2018 from `https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource`.

9. Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing Transparency Design into Practice. In *23rd International Conference on Intelligent User Interfaces (IUI '18)*. ACM, New York, NY, USA, 211–223. DOI: `http://dx.doi.org/10.1145/3172944.3172961`

10. Adrienne Porter Felt, Elizabeth Ha, Serge Egelman, Ariel Haney, Erika Chin, and David Wagner. 2012. Android Permissions: User Attention, Comprehension, and Behavior. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS '12)*. ACM, New York, NY, USA, Article 3, 14 pages. DOI: `http://dx.doi.org/10.1145/2335356.2335360`

11. Mireille Hildebrandt. 2016. The New Imbroglio – Living with Machine Algorithms. *The Art of Ethics in the Information Society* (2016).

12. Immanuel Kant. 2013. *An Answer to the Question: "What is Enlightenment?"*. Penguin UK.

13. Judy Kay and Bob Kummerfeld. 2013. Creating Personalized Systems That People Can Scrutinize and Control: Drivers, Principles and Experience. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 4, Article 24 (Jan. 2013), 42 pages. `DOI: http://dx.doi.org/10.1145/2395123.2395129`

14. Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2017. User Preferences for Hybrid Explanations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17)*. ACM, New York, NY, USA, 84–88. `DOI: http://dx.doi.org/10.1145/3109859.3109915`

15. Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. 2015. Principles of Explanatory Debugging to Personalize Interactive Machine Learning. In *Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15)*. ACM, New York, NY, USA, 126–137. `DOI: http://dx.doi.org/10.1145/2678025.2701399`

16. Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. 2012. Tell Me More?: The Effects of Mental Model Soundness on Personalizing an Intelligent Agent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1–10. `DOI: http://dx.doi.org/10.1145/2207676.2207678`

17. Todd Kulesza, Simone Stumpf, Margaret Burnett, Weng-Keen Wong, Yann Riche, Travis Moore, Ian Oberst, Amber Shinsel, and Kevin McIntosh. 2010. Explanatory Debugging: Supporting End-User Debugging of Machine-Learned Programs. In *Proceedings of the 2010 IEEE Symposium on Visual Languages and Human-Centric Computing (VLHCC '10)*. IEEE Computer Society, Washington, DC, USA, 41–48. `DOI: http://dx.doi.org/10.1109/VLHCC.2010.15`

18. Brian Y. Lim and Anind K. Dey. 2009. Assessing Demand for Intelligibility in Context-aware Applications. In *Proceedings of the 11th International Conference on Ubiquitous Computing (UbiComp '09)*. ACM, New York, NY, USA, 195–204. `DOI: http://dx.doi.org/10.1145/1620545.1620576`

19. Brian Y. Lim and Anind K. Dey. 2010. Toolkit to Support Intelligibility in Context-aware Applications. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp '10)*. ACM, New York, NY, USA, 13–22. `DOI: http://dx.doi.org/10.1145/1864349.1864353`

20. Bin Liu, Mads Schaarup Andersen, Florian Schaub, Hazim Almuhimedi, Shikun (Aerin) Zhang, Norman Sadeh, Yuvraj Agarwal, and Alessandro Acquisti. 2016. Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permissions. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX Association, Denver, CO, 27–41. `https://www.usenix.org/conference/soups2016/technical-sessions/presentation/liu`

21. Tim Miller. 2017. Explanation in Artificial Intelligence: Insights from the Social Sciences. *Computing Research Repository (CoRR)* abs/1706.07269 (2017). `http://arxiv.org/abs/1706.07269`

22. Christoph Molnar. 2018. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Retrieved February 7, 2018 from `https://christophm.github.io/interpretable-ml-book/`.

23. Deirdre K. Mulligan and Kenneth A. Bamberger. 2013. What Regulators Can Do to Advance Privacy Through Design. *Commun. ACM* 56, 11 (Nov. 2013), 20–22. DOI:http://dx.doi.org/10.1145/2527185

24. Patricia A. Norberg, Daniel R. Horne, and David A. Horne. 2007. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs* 41, 1 (2007), 100–126. DOI: http://dx.doi.org/10.1111/j.1745-6606.2006.00070.x

25. Pearl Pu and Li Chen. 2006. Trust Building with Explanation Interfaces. In *Proceedings of the 11th International Conference on Intelligent User Interfaces (IUI '06)*. ACM, New York, NY, USA, 93–100. DOI: http://dx.doi.org/10.1145/1111449.1111475

26. Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM, New York, NY, USA, 1135–1144. DOI: http://dx.doi.org/10.1145/2939672.2939778

27. Irina Shklovski, Scott D. Mainwaring, Halla Hrund Skúladóttir, and Höskuldur Borgthorsson. 2014. Leakiness and Creepiness in App Space: Perceptions of Privacy and Mobile App Use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, 2347–2356.

28. Anna Cinzia Squicciarini, Mohamed Shehab, and Federica Paci. 2009. Collective Privacy Management in Social Networks. In *Proceedings of the 18th International Conference on World Wide Web (WWW '09)*. ACM, New York, NY, USA, 521–530. DOI: http://dx.doi.org/10.1145/1526709.1526780

29. Jennifer Tidwell. 2010. *Designing Interfaces: Patterns for Effective Interaction Design.* (2 ed.). O'Reilly Media.

30. Joe Tullio, Anind K. Dey, Jason Chalecki, and James Fogarty. 2007. How it Works: A Field Study of Non-technical Users Interacting with an Intelligent System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, 31–40.

31. Alfredo Vellido, José David Martín-Guerrero, and Paulo J. G. Lisboa. 2012. Making Machine Learning Models Interpretable. In *ESANN*, Vol. 12. Citeseer, 163–172.

32. Sandra Wachter, Brent Mittelstadt, and Luciano Floridi. 2017. Transparent, Explainable, and Accountable AI for Robotics. *Science Robotics* 2, 6 (May 2017). http://dx.doi.org/10.1126/scirobotics.aan6080

33. Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. 2014. A Field Trial of Privacy Nudges for Facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2367–2376. DOI: http://dx.doi.org/10.1145/2556288.2557413