SN Social Sciences (2025) 5:99
https://doi.org/10.1007/s43545-025-01129-2

ORIGINAL ARTICLE

Addressing the potential social risk of self-reported data within a computational-intensive world

Salvatore Flavio Pileggi[1]

Received: 23 October 2024 / Accepted: 30 June 2025 / Published online: 14 July 2025

Abstract

This paper addresses the potential social risk associated with the capability to infer sensitive opinions from large self-reported data within a computational-intensive world, in which AI is pervasively and inherently adopted as part of the resulting socio-technical system. Such a social risk should be framed assuming a variety of socio-political contexts, including also non-democratic systems or, more in general, systems with significant lacks in terms of human rights. A simplified view of social risk is considered proportional to the sensitivity of the information and the prediction performance. The related computational experiments are conducted by applying Machine Learning techniques (Neural Networks) on a pre-existent case study based on a subset of the popular World Values Survey. Despite such a use case is not explicitly designed to maximise the prediction performance and is characterised by low dimensionality, the empirical results pointed out an overall interesting capability to infer potentially sensitive information. Additionally, the prediction accuracy resulted to be proportional to the likelihood of data to change along the time. Those results are discussed in context in the paper, looking holistically at the associated social risk, as well as at possible practical implications. In a continuously evolving context, characterised by fast advances of AI technology in contrast with a lack of systematic frameworks for reasoning about risk, uncertainty, and their potentially catastrophic consequences, this study focuses on computational experimentation and case studies to further stimulate the convergence of analysis frameworks and to nurture awareness from both a social and a user perspective.

Keywords Social risk · Ethical AI · Socially responsible AI · Privacy · Risk awareness

✉ Salvatore Flavio Pileggi
SalvatoreFlavio.Pileggi@uts.edu.au

[1] University of Technology Sydney, 15 Broadway, Ultimo NSW 2007, Australia

Introduction

More and more scientists from different disciplines are sending out explicit calls to take AI risks seriously (e.g (Novelli et al. 2023)). These concerns are well-reflected in the numerous attempts to define ethical principles for trustworthy AI, as well as in some political initiatives and legislations, such as the EU Artificial Intelligence Act (AIA) (2020/1828), which lays the foundations for the regulation of AI as a disruptive technology in the EU. To remark that most frameworks explicitly or implicitly assume an underlying context of democracy, freedom, inclusion and respect of fundamental human rights. Additionally, the perception of risk in a broader context may significantly vary, especially when considered looking at the inherently coupled relationship risk/benefit (Bao et al. 2022).

The different social implications are more or less directly addressed in a huge number of contributions in literature, which address at a different level of detail the inherent trade-off between opportunity and risk. For instance, in (Raso et al. 2018–6) the focus is explicitly on human rights, while (Panch et al. 2019) considers public health, as well as more generic considerations are provided in (Paaß and Hecker 2023). In general terms, the active research on ethical aspects and related practices aims to progressively define and foster a safer principled framework (Floridi and Cowls 2022; Floridi et al. 2018) and, therefore, a path to sustainable (Van Wynsberghe 2021) and socially responsible (Cheng et al. 2021) AI, under the intrinsic assumption that principles alone cannot guarantee ethical AI (Mittelstadt 2019).

This paper has a much more narrowed and specific focus on the social dimension. It contributes to demonstrate the potential social risk associated with large self-reported data within a computational-intensive world, in which AI is pervasively and inherently adopted as part of the resulting socio-technical system, by empirical assessment. Such risks should be framed, analysed and properly understood according to a realistic contextual approach, which considers democratic and non-democratic countries, as well as complex socio-political situations (De Mesquita et al. 2005) in which human rights are not always fully acknowledged and respected, resulting overall in a significant diversity in terms of constitutional, democratic and social principles.

The theoretical and practical relevance of privacy is a well consolidated concept (Rachels 1975), characterised by a number of social and political dimensions (Westin 2003). However, the concept of "privacy" is fairly elusive (Elliott and Soifer 2022) and not always well understood, measured, and reduced within AI systems (Curzon et al. 2021). An increasingly data and computational intensive society has defined a new reality in which privacy needs to be considered within the digital world (Regan 2002) with additional challenging issues (Zarsky 2019). Looking at self-reported data, the social risk associated is typically related, among others,to privacy, potentially identifiable data and an use of data for a purpose different from the original one. For instance, privacy issues are extremely common in online social networks (Zhang et al. 2010), where users may more or less intentionally disclose potentially sensitive information about themselves or others. Sensitive information is a broad concept and, indeed, concrete definition can significantly vary from case to case. Sensitivity is mostly associated with, but not limited to, race/ethnic origin, socio-political activity,

religious beliefs, sexual orientation, health and criminal record. Opinions, perceptions and believes are specifically object of study in this work.

More in general, minor or major concerns rise with the application of cutting edge technology that implies some social risk. For example, it's the case of AI for law enforcement, whose adoption is still pending over the resolution of critical issues (Raaijmakers 2019) in addition to intrinsic risks (e.g. preventive justice). Among the risks, there is also the potential capability to infer opinions about sensitive matters by adopting sophisticated computational techniques on large amounts of data. Concrete examples are opinions not well accepted, or even outlawed, in certain countries (e.g. homosexuality) and political positions, such as on conflicts, with legal consequences in some cases. A more systematic approach to establish predictive models might foster social engineering, understood as efforts or actions to influence attitudes and social behaviour on a large scale, as well as questionable social artefacts, such as social scoring. The former is not always in line with individual freedom, while the latter is object of controversy.

In the specific context of this paper, social risk is informally defined as a direct or indirect risk of damage or harm, even on a large scale. More concretely, this paper addresses the potential risk associated with the capability to infer sensitive opinions from large self-reported data in light of the recent advances of AI. In the adopted model, the social risk is considered proportional to the relationship between the sensitivity of the information and the prediction performance. The related computational experiments are conducted by applying Machine Learning techniques on a pre-existent case study, which is not explicitly designed to maximise the prediction performance and is characterised by low dimensionality.

Given a lack of systematic frameworks for reasoning about risk, uncertainty, and their potentially catastrophic consequences (Zhang et al. 2022), this works contributes to enhance risk awareness though an evidence-based approach supported by empirical assessment. Indeed, we believe that awareness and case studies are a determinant to promote and foster socially-responsible AI technology. That is in line with recently identified gaps. For instance, looking at the AI-driven evolution of Smart City (Yigitcanlar et al. 2020), there is an explicit reference to a generic limited scholarly research investigating the risks and consequent potential disruptions of a wide AI utilization. The focus of this study on computational experimentation and case studies to enable a more comprehensive analysis explicitly aims to (i) further stimulate the convergence of analysis frameworks towards a risk-aware model and (ii) nurture awareness from a user perspective, as well as from a more generic social perspective as part of a constantly evolving socio-technical context.

Structure of the paper

The paper is arranged according to a classic structure which includes a brief description of the key methodological aspects in Sec. 2, followed by an overview of the computational results and related discussion (Sec. 3 and 4 respectively).

Methodology and approach

As previously discussed, from a methodological perspective, the most significant challenge is related to the establishment of a research framework that allows a consistent critical analysis of the topic in the context of its inherent complexity within the current technological landscape.

AI systems are indeed quickly and consistently evolving in terms of capability and autonomy towards General AI, in contrast with AI safety research, which is much slower and, consequently, lagging in a context of lack of consensus at a strategic and operational level (Bengio et al. 2024). The different rising governance initiatives lack de facto the mechanisms and the institutions to prevent misuse and recklessness (Bengio et al. 2024).

This work proposes a simple case of misuse through an empirical assessment of a relevant case study. The integrated approach is based on a specific model for the kind of social risk object of analysis. Details about the case study and the model are briefly discussed in this section. A description/definition of symbols and acronyms is reported in Table 1.

Modelling the social risk associated with self-reported data

The social risk model underlying this work is reported in Fig. 1. Given a dataset S of $n \in N$ variables and a categorization C computed from those variables ($C = f(n)$), a predictor P aims to predict the category $c \in C$, taking in input a subset of $m \in M$ variables ($M \subset N$). Therefore, the set $K = N - M$ includes those variables that are implicitly part of the categorization but are not used as an input for the predictor. The implicit variables (K) can be inferred from the category (C).

In the context of this work, S is produced from self-reported data, typically survey data or equivalent. Additionally, its size is supposed to be suitable to machine-learning approaches, predictive models in this specific case.

The potential social risk R of the prediction is function of the sensitivity $w_i, i \in K$ associated with the set of implicit variables K and with the predictor's performance. The former factor reflects the simple empirical assumption that variables may have

Table 1 Description/Definition of symbols and acronyms

Symbol/Acronym	Definition/Description
S	Available dataset (self-reported data/survey data)
N	Dimensionality of S
C	Set of categories
M	Subset of variables ($M \subset N$) adopted for training.
K	Implicit Variables ($N - M$)
P	Predictive Model
W	(Topic) Sensitivity associated with K
R	Social Risk associated with P
Exp.#	Experiment
Dim(J)	Dimensionality of a given set J

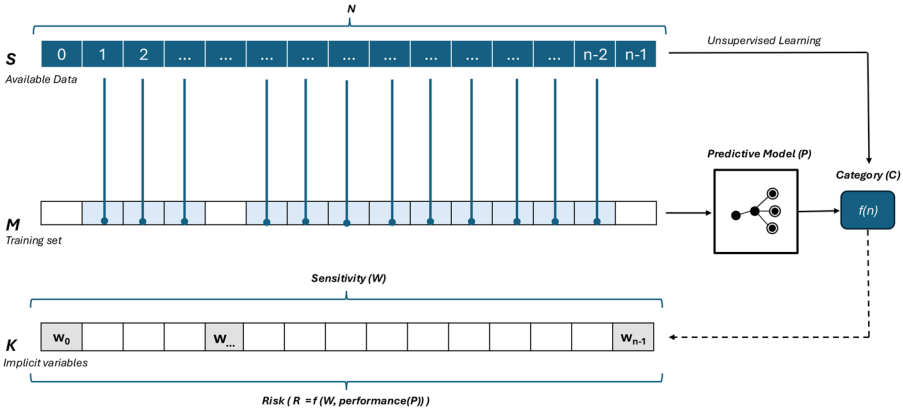

Fig. 1 Social risk model in the context of the experiment performed

a very different contextual sensitivity, while the latter realistically assumes that a higher capability of prediction may increase the related social risk.

Case study: values, opinions and perceptions

The study reported in this paper is performed on the *World Values Survey (WVS)* (JD Systems Institute & WVSA 2022), which has been extensively adopted by the research community to undertake a large number of studies in social sciences, as well as in other disciplines. The conceptualization underlying this work has been defined in a previous study (Pileggi 2024), where a small subset of WVS has been identified to define a number of profiles through unsupervised learning techniques.

The considered factors and the resulting profiles are represented in Fig. 2. The representation maintains the original scale with values in the range 1 (maximum relevance) and 4 (minimum relevance). In general terms, factors are assessed by considering one or more variables (or attributes) - i.e. questions in surveys. In this specific case study, there is a one-to-one correspondence between factors and variables. The 16 considered variables are classified in three different categories (*values*, *opinions* and *perceptions*) according to their likelihood to change along the time (Pileggi 2024). Each variable is reported with the original ID (question #) in WVS.

The set S includes the whole set of variables and can be formally defined like the composition of three disjoint sub-set of variables ($S = \{values \cup opinions \cup perceptions\}$).

The profiles provide a good overview of the relationships among the different factors, as well as insight about the value of predictions. In the original study there is no formal analysis of the potential factor sensitivity. Although an effective analysis should be contextual, in general terms and in the specific context of this work, the following variables are considered to be potentially sensitive:

- *Gender discriminatory opinions (Q29)*. WVS data has shown a diversity of answers to the related question. It reflects strong underlying cultural differences but also inequality (Ridgeway 2011), violence (Krahé 2016), as well as lack of

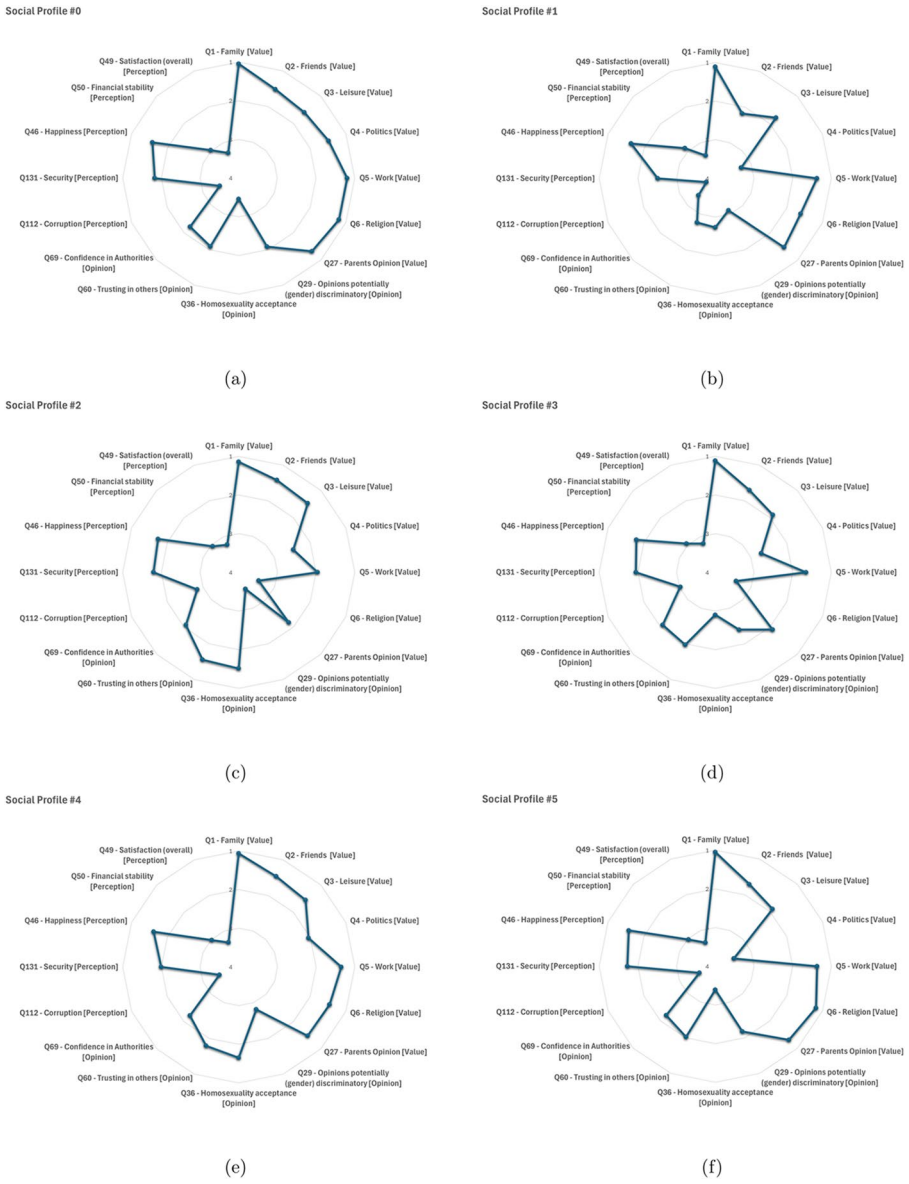

Fig. 2 Conceptualization and profiles from (Pileggi 2024). The representation adopts the original scale (numerical value between 1 and 4), where the higher relevance of a factor is associated with the lower value in such a scale - i.e. 1 is the maximum relevance and 4 is the minimum relevance

rights (Sullivan 1994). While the profiles built in (Pileggi 2024) from WVS data are holistic and do not take into account cultural backgrounds or similar aspects, the perception of this factor may be considered sensitive in certain multi-cultural contexts.

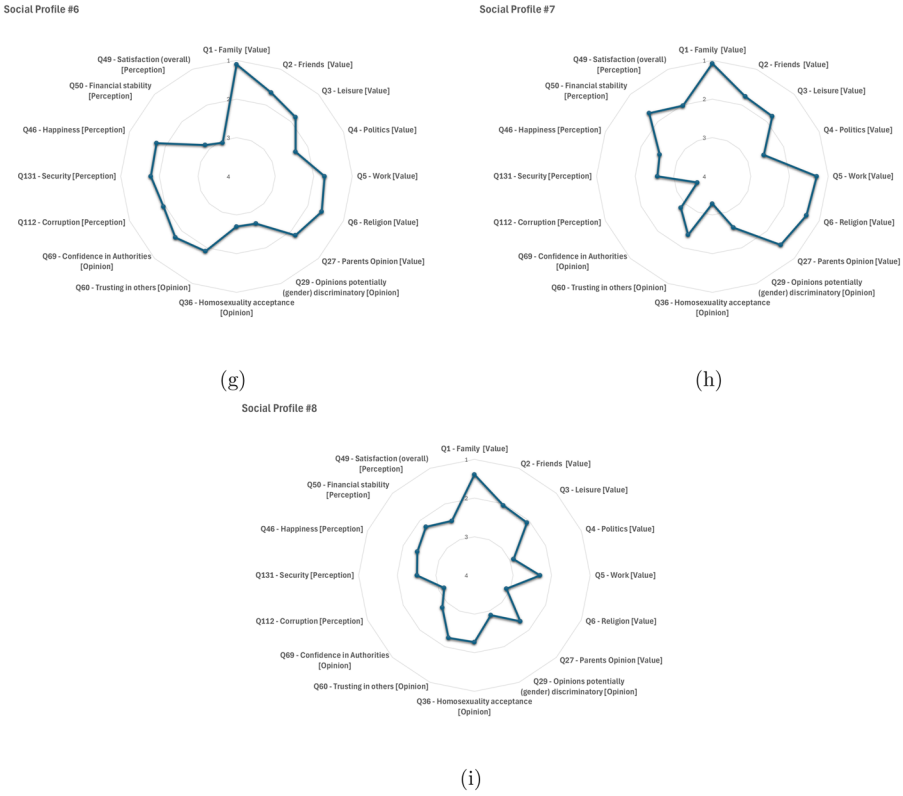

Fig. 2 (continued)

- *Homosexuality acceptance (Q36)*. Homosexuality is perceived in a different way (Kite and Bryant-Lees 2016), even within cultural contexts that formally accept it (e.g. (Patacchini et al. 2015)). Discrimination, homophobia and violence are well-known issues (Bernat et al. 2001), and in many countries there are still a lack of rights (Encarnación 2014). Noteworthy, despite a decriminalisation trend, homosexuality is currently illegal in 60+ countries. Penalties range in severity including capital punishment which is still accepted in certain countries (ILGA World Database).
- *Confidence in authorities (Q69) and Perception of corruption (Q112)*. This kind of factor may be very sensitive within non-democratic countries, including hybrid and authoritarian regimes (Our World in Data), as well as in pseudo-democracies where, despite self-claims, there is a fundamental lack of freedom. For instance, negative perceptions may be associated with a lack of support or alignment with the government/regime.

Computations

This section is logically structured in two different parts that address, respectively, the description of the adopted computational environment and the overview of results.

Computational framework

The computations have been performed by using algorithms implemented by the Scikit-learn Python Package (Pedregosa et al. 2011) as follows:

- *KMeans* (Sinaga and Yang 2020) has been adopted to define the different categories, re-proposing the approach in (Pileggi 2024).
- *Elbow algorithm* (Thorndike 1953) has provided a heuristic estimation of the optimal number of clusters as an input to the KMeans algorithm. As in the original conceptualization (Pileggi 2024), 9 clusters have been considered.
- *MLPClassifier* (Rosenblatt 1958); Hinton 1990) has supported the implementation of the predictive model. The training set includes 58248 lines, corresponding to the 80% of the total, while the testing set encompasses the remaining 20% (14562 lines). Additionally, the computations assume scaling and re-sampling, meaning the training set and the testing set have been sampled more than one time to perform independent experiments. Therefore, the scores reported below are the average on 3 samples. In terms of parameters, the solver is *lbfgs*, an optimizer in the family of quasi-Newton methods, while the neural network is configured to have one layer with 12 neurons.

The case study analysis consists of an empirical assessment based on a set of computational experiments. Experiments differ from each other because of their prediction goal, namely a given block of variables or individual variables, and result from different combinations of parameters.

Results

The summary of the conducted experiments is reported in Table 2. For each experiment, it is specified the training set M and the set of implicit variables K, as well as their dimensionality (*Dim(M* and *Dim(K)* respectively) and the related average prediction score computed on multiple samples.

The first experiment (*Exp.#1* in the table) aims at category validation as it assumes all available data as a training set ($M = S$). While it has no concrete meaning in terms of application because there is no implicit variable ($K = \varnothing$), the high prediction score (0.998) confirms the consistency of the considered categories.

Block prediction (*Exp.#2.1*, *Exp.#2.2* and *Exp.#2.3*) assumes two of the three predefined blocks as a training set ($M = \{m_i, m_j\}$ and $K = \{m_k\}$, where $i \neq j \neq k$ and $i, j, k \in \{values, opinions, perceptions\}$). For these experiments, the score ranges from 0.561 to 0.715, depending on the configuration of the training set. For block prediction, the training set is dimensionally smaller than in the previous experiment. It results in a much lower accuracy in this case. To note that the accuracy of the predictor follows a pattern that is proportional to the likelihood to change along the time as $Score(K = values) < Score(K = opinions) < Score(K = perceptions)$. Such a finding can be considered consistent as it is not completely proportional to the dimensionality of the training set (M=9, M=12, M=11 for *Exp.#2.3*, *Exp.#2.2* and

Table 2 Experiment details and results

Exp.#	Description	M	K	Dim(M)	Dim(K)	Score
1	Category Validation	S = {Q1, Q2, Q3, Q4, Q5, Q6, Q27, Q29, Q36, Q60, Q69, Q112, Q131, Q46, Q50, Q49}	∅	16	0	*0.998*
2.1	Block Prediction	values = {Q1, Q2, Q3, Q4, Q5, Q6, Q27}, opinions = {Q29, Q36, Q60, Q69}	perceptions = {Q112, Q131, Q46, Q50, Q49}	11	5	*0.715*
2.2	Block Prediction	values = {Q1, Q2, Q3, Q4, Q5, Q6, Q27}, perceptions = {Q112, Q131, Q46, Q50, Q49}	opinions = {Q29, Q36, Q60, Q69}	12	4	*0.650*
2.3	Block Prediction	opinions = {Q29, Q36, Q60, Q69}, perceptions = {Q112, Q131, Q46, Q50, Q49}	values = {Q1, Q2, Q3, Q4, Q5, Q6, Q27}	9	7	*0.561*
3.0	Sensitive Variable(s) Prediction	Q1, Q2, Q3, Q4, Q5, Q6, Q27, Q60, Q131, Q46, Q50, Q49	Q29, Q36, Q69, Q112	12	4	*0.615*
3.1	Sensitive Variable(s) Prediction	Q1, Q2, Q3, Q4, Q5, Q6, Q27, Q60, Q112, Q131, Q46, Q50, Q49	Q29, Q36, Q69	13	3	*0.672*
3.2	Sensitive Variable(s) Prediction	Q1, Q2, Q3, Q4, Q5, Q6, Q27, Q60, Q69, Q131, Q46, Q50, Q49	Q29, Q36, Q112	13	3	*0.657*
4.1	Sensitive Variable(s) Prediction	Q1, Q2, Q3, Q4, Q5, Q6, Q27, Q36, Q60, Q69, Q112, Q131, Q46, Q50, Q49	Q29	15	1	*0.887*
4.2	Sensitive Variable(s) Prediction	Q1, Q2, Q3, Q4, Q5, Q6, Q27, Q29, Q60, Q69, Q112, Q131, Q46, Q50, Q49	Q36	15	1	*0.792*
4.3	Sensitive Variable(s) Prediction	Q1, Q2, Q3, Q4, Q5, Q6, Q27, Q29, Q36, Q60, Q112, Q131, Q46, Q50, Q49	Q69	15	1	*0.880*
4.4	Sensitive Variable(s) Prediction	Q1, Q2, Q3, Q4, Q5, Q6, Q27, Q29, Q36, Q60, Q69, Q131, Q46, Q50, Q49	Q112	15	1	*0.880*

Exp.#2.1 respectively). However, the low performance in *Exp.#2.3* may have been affected by the lower dimensionality of the training set.

Similarly, *Exp.#3.0* assumes a customised set of implicit variables, which is composed of the most sensitive ones ($K = \{Q29, Q36, Q69, Q112\}$). The score is 0.615, namely a value in between the scores of *Exp.#2.2* and *Exp.#2.3*. Assuming some semantic correlation between *Q69* and *Q112*, the training set dimensionality can be increased to define *Exp.#3.1* and *Exp.#3.2*, whose score is a higher than in *Exp.#3.0*.

The last set of experiments (*Exp.#4.1–4.4*) assumes one single sensitive feature as an implicit variable. It results in significantly higher performance, with a score in a range 0.792–0.887.

A holistic view of results is proposed in Fig. 3, where the scores are represented as a function of the training set dimensionality.

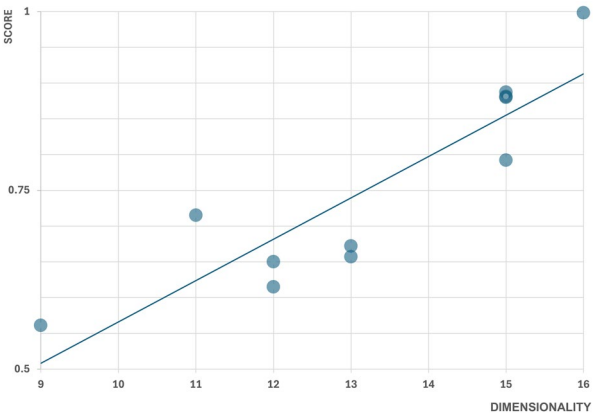

Fig. 3 Score as a function of the training set dimensionality

Discussion

As previously mentioned, the conducted experiment is underpinned by generic data from the research community and refers to an existent case study that has not been designed to maximise predictions. Apart from being characterised by a very limited dimensionality, such a case study doesn't consider any demographic data. A very first reflection is, therefore, on the nature of data. Indeed, in the modern technological landscape and data-intensive society, high-dimensionality dataset, including also demographic data, are relatively common. The *World Values Surveys* (JD Systems Institute & WVSA 2022) is an example. To note that datasets capturing or reflecting personal values, opinions, believes and perceptions may be produced with methods other than surveys, from Social Networks for instance. Similarly, profiles may be built from multiple datasets by adopting more complex techniques.

The patterns related to block prediction (Table 2) seem to be consistent with the assumptions in the original case study (Pileggi 2024), where the different attributes are classified based on their likelihood to potentially change along the time. Unsurprisingly, the score associated with the prediction of the most volatile block of attributes (perceptions) is the higher with a progressive decreasing pattern as a function of the theoretical volatility. Opinions and perceptions (as defined in the original case study) are probably more critical in terms of potential social risk. This consideration can be reasonably generalised to a certain degree, although sensitive variables may potentially belong to any category.

An additional discussion point is about uncertainty. Indeed, self-reported data may often include a number of incomplete data points, as well as complete answers that show a preference for not answering to a given question (typically the option "prefer not to say"). While in computational terms such situations determine a certain degree of uncertainty, from a semantic perspective unanswered questions might be an indicator of sensitivity. Missing data can result from different situations, such as unclear questions, so generalizations are evidently not possible. However, the relationship between missing data and qualitative parameters - i.e. sensitivity and criticality - deserves additional investigation and empirical assessment.

Conclusions and future work

In a context of intense discussion within the research community on the multifaceted implications of AI from a social perspective, this paper presents a narrowed and specific focus on self-reported (or semantically equivalent) data by contributing to demonstrate potential social risks through empirical assessment.

The experiment conducted applies relatively simple computational techniques to predict potentially sensitive attributes from other attributes. While statistical correlations are inherent in the underlying dataset and may be exploited in the full respect of ethical principles and human values, misuse and unethical applications on a large scale constitute a concrete potential social risk within the modern data-intensive society. This is a critical aspect looking at the actual socio-political complexity of the different countries. Such a complexity not always allows a full alignment with the assumptions underlying the frameworks for ethical and trustworthy AI.

Misuse of data and technology is definitely not a novelty. However, it may become more critical in the era of AI, which enables in fact unprecedented capabilities. Therefore, such a study wants to further foster social awareness and can be understood as an additional call for a more specific socio-technical approach when dealing with disruptive technology. In the specific case of AI, the direct or indirect social implications may be significant, even in an apparently harmless context such us the anonymous data addressed in the paper.

More holistically, this work contributes to progressively build social awareness by providing valuable examples of potential risk on a broad scale. Such a value is evidently not alternative but rather complementary to the principled approach to ethics for technology use. Indeed, more and more often the general principles are associated with real world scenarios and applications to enhance their understanding in context and to better frame the different direct or indirect implications for individuals, organizations and society. On the other side, case studies enable a feedback process to refine and further elaborate principles and requirements, characterised by an inherent high level of abstraction due to their genericness. The specific case study on self-reported data addressed in the paper can be critical within a digitalised society, where computation capabilities are constantly growing as part of an evolving and increasingly sophisticated technology. The capability of inference by establishing statistical correlations is unprecedented and becomes critical when sensitive topics are involved in a context of intrinsic socio-political complexity. The experiment described in the paper, based on the popular dataset *World Values Survey* and relatively simple Machine Learning techniques, demonstrates a potential ability of profiling individuals and predict opinions/believes on sensitive topics from minimal input. Most (if not all) ethical frameworks explicitly refer to privacy protection in a context of full respect of human rights and values. However, inherent risks should be considered looking at a variety of possible situations and contexts in which a wrong use of technology may happen.

Future work will aim to further investigate potential social risks though an evidence-based approach. It will include an extended experimentation at a larger scale adopting more sophisticated computational methods to maximise predictions on all available data. Last but not least, additional research is needed to better understand how ethical principles may be understood, interpreted and operationalised in the different contexts.

Acknowledgements I would like to thank the team behind the World Value Survey for making the dataset freely available for the community, as well as the anonymous reviewers, who have provided extensive and constructive feedback.

Author contributions Not Applicable.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. Open Access funding enabled and organized by CAUL and its Member Institutions. This research was not externally funded.

Data availability This work is based exclusively on secondary data. The original dataset, the *World Values Survey* (JD Systems Institute & WVSA 2022), is cited in the paper and is available to the community.

Declarations

Ethical approval Not applicable.

Informed consent Not applicable.

Conflict of interest The author declares that he has no competing interests.

References

Bao L, Krause NM, Calice MN, Scheufele DA, Wirz CD, Brossard D, Newman TP, Xenos MA (2022) Whose AI? How different publics think about AI and its social impacts. Comput In Hum Behav 130(107182)

Bengio Y, Hinton G, Yao A, Song D, Abbeel P, Darrell T, Harari YN, Zhang Y-Q, Xue L, Shalev-Shwartz S et al. (2024) Managing extreme AI risks amid rapid progress. Science 384(6698):842–845

Bernat JA, Calhoun KS, Adams HE, Zeichner A (2001) Homophobia and physical aggression toward homosexual and heterosexual individuals. J Abnormal Phychol 110(1):179

Cheng L, Varshney KR, Liu H (2021) Socially responsible ai algorithms: issues, purposes, and challenges. J Artif Intell Res 71:1137–1181

Curzon J, Kosa TA, Akalu R, and El-Khatib K (2021) Privacy and artificial intelligence. IEEE Trans Artif Intell 2(2):96–108

De Mesquita BB, Downs GW, Smith A, Cherif FM (2005) Thinking inside the box: a closer look at democracy and human rights. Int Stud Q 49(3):439–457

Elliott D, Soifer E (2022) AI technologies, privacy, and security. Front Artif Intell 5(826737)

Encarnación OG (2014) Gay rights: why democracy matters. J Democr 25(3):90–104

European Union (2020/1828) Regulation (eu) 2024/1689 of the European parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence and amending regulations (ec) no 300/2008, (eu) no 167/2013, (eu) no 168/2013, (eu) 2018/858, (eu) 2018/1139 and (eu) 2019/2144 and directives 2014/90/eu, (eu) 2016/797 and (eu). http://data.europa.eu/eli/reg/2024/1689/oj. 18 September 2024

Floridi L, Cowls J (2022) A unified framework of five principles for ai in society. Machine Learning And The City: Applications In Architecture And Urban Design 535–545
Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, Luetge C, Madelin R, Pagallo U, Rossi F et al. (2018) Ai4people—an ethical framework for a good ai society: opportunities, risks, principles, and recommendations. Minds Machines 28:689–707
Hinton GE (1990) Connectionist learning procedures. In: Machine learning. Elsevier, pp 555–610
ILGA World Database Legal frameworks - criminalisation of consensual same-sex sexual acts. https://database.ilga.org/criminalisation-consensual-same-sex-sexual-acts. 10 September 2024
JD Systems Institute & WVSA (2022) European values study and world values survey: joint EVS/WVS 2017–2022 dataset (joint EVS/WVS). https://doi.org/10.14281/18241.21. Dataset Version 4.0.0
Kite ME, Bryant-Lees KB (2016) Historical and contemporary attitudes toward homosexuality. Teach Psychol 43(2):164–170
Krahé B (2016) Violence against women. Aggression And Violence 251–268
Mittelstadt B (2019) Principles alone cannot guarantee ethical ai. Nat Mach Intell 1(11):501–507
Novelli C, Casolari F, Rotolo A, Taddeo M, Floridi L (2023) Taking ai risks seriously: a new assessment model for the ai act. AI Soc 1–5
Our World in Data Countries that are democracies and non-democracies. https://ourworldindata.org/grapher/countries-democracies-nondemocracies-eiu. 10 September 2024
Paaß G, Hecker D (2023) Ai and its opportunities, challenges and risks. Artificial Intelligence: What Is Behind The Technology Of The Future? 363–428
Panch T, Pearson-Stuttard J, Greaves F, Atun R (2019) Artificial intelligence: opportunities and risks for public health. Lancet Digit Health 1(1):e13–e14
Patacchini E, Ragusa G, Zenou Y (2015) Unexplored dimensions of discrimination in europe: homosexuality and physical appearance. J Popul Econ 28:1045–1073
Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al. (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830
Pileggi SF (2024) A hybrid approach to analysing large scale surveys: individual values, opinions and perceptions. SN Soc Sci 4(8):144
Raaijmakers S (2019) Artificial intelligence for law enforcement: challenges and opportunities. IEEE Secur Priv 17(5):74–77
Rachels J (1975) Why privacy is important. Philosophy & Public Affairs 323–333
Raso FA, Hilligoss H, Krishnamurthy V, Bavitz C, Kim L (2018-6, 2018) Artificial intelligence & human rights: opportunities & risks. Berkman Klein Center Research Publication
Regan PM (2002) Privacy as a common good in the digital world. Information, Communication & Society 5(3):382–405
Ridgeway CL (2011) Framed by gender: how gender inequality persists in the modern world. Oxford University Press
Rosenblatt F (2021). The perceptron: a probabilistic model for information storage and organization 1958
Sinaga KP, Yang M-S (2020) Unsupervised k-means clustering algorithm. IEEE Access. 8:80716–80727
Sullivan DJ (1994) Women's human rights and the 1993 world conference on human rights. Am J Int Law 88(1):152–167
Thorndike RL (1953) Who belongs in the family? Psychometrika 18(4):267–276
Van Wynsberghe A (2021) Sustainable ai: ai for sustainability and the sustainability of ai. AI Ethics 1(3):213–218
Westin AF (2003) Social and political dimensions of privacy. Journal Of Social Issues 59(2):431–453
Yigitcanlar T, Desouza KC, Butler L, Roozkhosh F (2020) Contributions and risks of artificial intelligence (ai) in building smarter cities: insights from a systematic review of the literature. Energies 13(6):1473
Zarsky TZ (2019) Privacy and manipulation in the digital age. Theoretical Inquiries In Law 20(1):157–188
Zhang C, Sun J, Zhu X, Fang Y (2010) Privacy and security for online social networks: challenges and opportunities. IEEE Network 24(4):13–18
Zhang X, Chan FT, Yan C, Bose I (2022) Towards risk-aware artificial intelligence and machine learning systems: an overview. Decis Support Syst 159(113800)