23rd International Conference on Harbor, Maritime and Multimodal Logistic Modeling & Simulation
18th International Multidisciplinary Modeling & Simulation Multiconference

ISSN 2724-0339 ISBN 978-88-85741-59-1 © 2021 The Authors.
doi: 10.46354/i3m.2021.hms.011

Data Value as the basis for Human Behavior Modeling

Agostino G. Bruzzone[1,*], Francesco Longo[2], Giulio Franzinetti[3], Alberto De Paoli[4], Enrico Ferrari[4]

[1] Simulation Team, SIM4Future, via Trento 43, 16145 Genova, Italy
[2] MSC-LES, DIMEG, University of Calabria, Via Pietro Bucci, Cubo 45 C, 87036 Rende
[3] Lio-Tech ltd., London, UK
[4] Simulation Team, Italy

*Corresponding author. Email address: agostino.bruzzone@simulationteam.com

Abstract

This paper introduces the importance of creating models to evaluate the value of data regarding their potential to extract information to identify human behaviors, attitudes, and characteristics. This is just a preliminary overview on this potential and consider that in the future, these values could become parts of the assets of companies if properly acquired and processed in order to respect all regulations and all rights of all the parties.

Keywords: Data Value, Financial Activities, Video Processing, Human Behavior Modeling

1. Introduction

The current times are defined as the Age of Information; therefore, in this first decades of the third millennium, we are dealing primarily with data: big data extracted in digital forms directly by sensor grid, heterogeneous networks, media, socials, IoT/IIoT (Internet of Things, Industrial Internet of Things), Digitalization of Companies & Society, etc.

Data by themselves are often full of inconsistencies, the sets are usually incomplete, and their volatility could lead to quick obsolescence; therefore, our capability to process them by new intelligent systems thanks to advances in computational power, connectivity, and interoperability is an order of magnitude higher than 20 years ago, providing us the possibility to transform these Data in quantifiable value (Bughin et al.,2010). The authors in this paper introduce the concept that these data could be evaluated as Economic Assets for the Companies based on their potential to create value. This value relies on the advanced use of innovative architectures based on Modeling, Simulation, Artificial Intelligence (AI) & Data Analytics. This approach is the basis for Strategic Engineering, the new fast-evolving discipline that aims to combine these techniques in a closed loop with actual data to support decisions (Bruzzone et al., 2018).

The adverse effects of Data in terms of economic value have been already extensively investigated concerning the cyber layers and their vulnerability (Poyraz et al., 2020); in fact, now it turns out that it is possible also to consider positively specific sets of data and evaluate them as tangible assets and resources that could increase the value of a company based on their potential (Cavanillas et al., 2016). This paper addresses these issues in respect of the finance sector.

2. Data as Asset with a Value such as a Mine

The concept of considering data as valuable assets is well known and even emerging as a significant concept in recent times (Leonelli 2019). Nowadays, this is the result of having the capability to process these big data by AI (Artificial Intelligence) to extract meaningful information corresponding to the actual value. The concepts on this aspect have been investigated even in the past. Therefore, the digitalization of the Company and Society is offering the possibility to access in a more straightforward, affordable, and quick way a large quantity of data, enabling possibilities that before were not sustainable. So, we can consider the value of this data similar to the value of a mine of iron ore, or gold ore, that could allow us to extract and process it, up to producing steel bars or pure ingots. The value of such data relies on the quality of the source and in our capability to discriminate between gold and iron, as well as on our capability to extract, refine and generate valuable information.

In this sense, it is evident that the economic quantification of these data is strongly related to their capability to be effectively used by artificial intelligence solutions to create value in terms of usable knowledge and support to critical decisions.

For instance, this paper considers financial intermediaries companies involved in the sale and management of investment products. Special attention in this paper is devoted to using data from video recordings of clients also used as a digital signature of the agreements, negotiations, and data about financial transactions and customer profiles over time and in different areas.

From this point of view, it is crucial to outline that the three significant kinds of data profiling are usually considered: structure discovery, content discovery, and relationship discovery (Bhardway & Long, 2021)

The structure discovery, or structure analysis, focuses on validating data in terms of consistency and format and relies on many different methods, including pattern matching; this analysis also obtains simple basic statistics over the data and their confidence bands. Vice versa, the content discovery is concentrated on a close look at the individual database elements to check data quality; for instance, missing data, null values, and incorrect numbers present in the database are essential to clean it and avoid adverse effects while feeding Machine Learning.

Also, relationship discovery is crucial because it involves the analysis of cross relations between the different data sets, usually by creating metadata analysis devoted to tracking critical relationships between data and identifying potential overlaps and misalignments.

In addition, it makes sense to point out that in finance and trading, one advantage could be obtained by adequately understanding the profile of the counterparts; such an understanding could be achieved by modeling human behavior in respect of decisions and negotiation. These models could be used anonymously to identify guidelines to focus on the most promising investors and to adapt at the same time the products to their needs. Obviously, at the same time, valuable data refers to the classification by behaviors, features, and attitudes in order to direct different marketing and business actions. Last but not least, it is evident that these data, in a more trivial way, are ready for carrying out individual profiling for multiple uses within and outside the financial sector. About this last aspect, there is complex evolution and onerous regulations on privacy around the globe that suggest high risks related to considering the exploitation of these aspects (Williams, 2017; Wimmer 2018; Mavriki & Karyda, 2018; Aridor et al., 2020; Aho & Duffield, 2020; Nicola & Pollicino, 2020). Therefore, despite the value of extracting information about his health, fashion preferences, interests from a video of a multi-millionaire, the terms of use of these data should be structured appropriately by experts. At the same time, potential critical changes could occur in the short term. Due to these reasons and considering that this aspect has been already investigated (Nemitz, 2018; Isaak & Hanna 2018; Ploug & Holm, 2021), in this paper, we do not consider too much this aspect, even if the extraction of this information could be essential to classify behaviors and attitude anonymously, creating generic profiles as well as symptoms that could lead to measure a favorable/adverse reactions respect investment types, proposals, and market evolution.

3. Human Behavior & Profiling Attitudes

Human behavior, even on fundamental decisions such as how to dress for a meeting, what App to use to address (e.g., urban car traffic, restaurants), is characterized by a personal touch, and it is not only an individual or a person, but corresponds even to his or her current emotional status, stress level, goals & aims (Pantic & Rothkrantz,2004; Achara et al., 2015; Sjöberg et al. 2016). The combination of a few of these elements often allows the identification of the main characteristics of a person with high validity. For instance, by aggregating the features extracted from smartphone data, it is possible to classify the Big-Five personality traits (extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience). Supervised learning can achieve quite good reliability (~75%) on these estimates (Chittaranjan et al., 2011).

Indeed there is solid literature regarding the use of different methods to finalize profiling of persons based on the texts and words, and it is possible to combine language use with personality traits and emotions and attitudes (Pennebaker et al., 2003); even the essential selection of words, verbs, prepositions, and determiners provides clues able to allow one to discriminate a person in terms of age, gender and condition (Schler et al., 2006; Rangel & Rosso 2016).

Nowadays, this analysis could be combined with speech recognition to create a compelling speaker profile in a video (Saha 2017).

We know, as humans, how to recognize emotions by facial expression; indeed, studies are dating back two decades for developing computer systems able to obtain high validity in this analysis of images based on facial muscle actions, and machine learning is further reinforcing this aspect (De Silva et al., 1997; Wingenbach et al., 2016, Abdulsalam et al., 2019)

There is exciting research about how negotiation and emotions could be crossed in order to support this process(Broekens et al., 2010); indeed, decision making is influenced not only by rational analysis but also by current status, stress, and emotions as well as by the background and attitude of the parties. In this sense, some potential models could develop a negotiation support system; obviously, it is fundamental to have valid data related to the specific subject and negotiation of interest (Kaya & Schoop, 2019).

This means that in the future, it could be strategic to develop data sets to train and update the correlations on emotional and rational negotiation to train intelligent systems as currently, we train people devoted to carrying out this task. In order to do that, it is crucial to identify the capability of some data set to support human behavior modeling in a broad sense (Maler et al.,2020).

Working on large data sets, it could be possible to extract essential indicators on behaviors from different sources; for instance, the history of individual data on smartphone use, in terms of number of calls, intensity over periods, durations of conversation, contact network, and mobility, allows for the classification of the socioeconomic status; in this way, it becomes possible to predict attributes for millions of people and their attitude to move and their expenses and investments (Blumenstock et al., 2015)

Therefore, it is interesting to outline that even in a limited data set, for instance, 26 couples living within a community for over one year, their social behavior could be evaluated in direct interactions by smartphones such as Bluetooth (face-to-face), messages calls, and messages phone class. These logs enable one to predict the behavior of these families with respect to expenses, in terms of propensity to explore the new business, customer fidelity, and overspending (Singh et al., 2013)

In addition, by using multiple data sources from video, it is possible to analyze it over time to extract information even if confidential information, for instance, estimating the flow of blood and how it fills the face as well as small motions that lead to understanding the stress level and emotional status of a subject (Sottilare & Proctor 2012; Wu et al. 2012). Indeed, today, it could be even considered possible to be able to carry out eye tracking to evaluate the reasoning and the attitude of a subject by using introductory video like that one provided by smartphones that are used during the financial transaction related to this case (Kafka et al., 2016)

4. Case Study

Based on the consideration mentioned above, there is an opportunity to valorize the data as new Financial Assets for Companies. Indeed, new opportunities could come to identify symptoms of a specific attitude or preference that could improve the finalization of deals.

Usually, these achievements are strongly related to the quantity of data and proper profiling of people; in this sense, the use of video during negotiation and deals extracting features and critical indicators based on observations cross-references with other individual data and boundary conditions could lead to creating value by Data Fusion & Machine Learning. This process should be able to fine-tune, dynamically Human Behavior Models to be used on only on the specific person but even to extract

As anticipated, the proposed case focuses on Financial Deals with investors that finalize actions from 100k USD up to a few Million.

This investment is related to Financial Products and results pretty fast being finalized, usually within less than an hour, potentially with multiple meetings; so, the capability to classify behaviors, attitudes, and identify symptoms of preferences relies on the promise of global, comprehensive, readily available data.

In principle, the combination of robust analytical methods with big data can support extraction of indicators to be used anonymously for being exploited too many potential users, reversing the usual formula to use an extensive data set for good profiling of a single individual and leading to the profiling of behavior models to be used on many deals by many users.

We mention that in our case, the data include videos of the deal; from this point of view, researchers confirm that subjects also engaged in basic economic games, pay great attention to the opponent's face (Rossi, Fasel & Sanfey, 2011). Indeed, in this case, the analysis demonstrated that the face of a decision-maker focusing on a deal contains valuable information for defining a winning strategy by modeling facial expressions of subjects and recording data over 60 participants, and extracting automatically key parameters temporal evolution (e.g., head pitch, yaw, and roll); the experimentation was aiming to predict the opponent's decision between offer and decision on the game and achieved an accuracy of 0.66 (chance = 0.50) in predicting decisions.

Laney and Crowley-Sweet respect Highways England consider that every 2 pounds of physical assets, the company could evaluate up to 1 pound of data assets.

Indeed, in order to create value from these data, it is required that the people understand their strategic value and manage them and the related ICT systems correctly (Crowley-Sweet 2021; Laney et al. 2021); similarly, the free access to personal web cameras on a laptop and related captured data for one month to be used freely has been evaluating equivalent to 1000 USD/month per person (Bloor, 2020). Recent researches have estimated a much more detailed set of tables related to data costs and reported in tables 1, 2, 3 (Steel, 2020). In general, it is crucial to define the procedure to evaluate the data value (Sengupta & Rusli, 2012; Bataineh et al., 2016; Spiekermann et al. 2017;

Age Range	Cost for Data [USD/Person]
18-24	0.36
25-35	0.11
35-44	0.12
45-54	0.27
>54	0.05

Table 1. Data Cost vs. Age

Ethnics Group	Cost for Data [USD/Person]
Caucasian	0.19
AfroAmerican	0.57
Hispanic	0.01
Asian	0.05
Native American	0.09

Table 2. Data Cost vs. Ethnicity

Family Income [USD/Year]	Cost for Data [USD/Person]	Family Income [USD/Year]	Cost for Data [USD/Person]
<10000	0.1	60000-69999	0.05
10000-19999	0.03	70000-79999	0.05
20000-29999	0.04	80000-99999	0.05
30000-39999	0.07	100000-119000	0.04
40000-49999	0.02	120000-149000	0.33
50000-59999	0.03	>149000	0.22

Table 3. Data Cost vs. Incomes

Investors have specific Attitudes on Risks, Business, and Behaviors we can perceive in terms of voice, tone, movements, expression, words to be cross related by using Machine Learning over the indicators extracted from video, audio, personal profile, synchronized notes by experts, history of transactions/deals and specific ongoing negotiation as proposed in figure 1. In these cases, it is usual to evaluate available data by fuzzification to extract the values of membership functions respect the indicators devoted to estimating the different parameters (Mandryk & Atkins, 2007; Chakraborty et al., 2009, Abdelhedi et al., 2016).

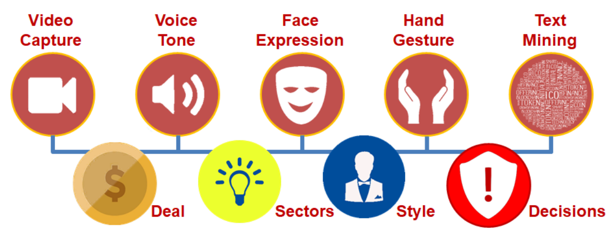

Figure 1. Data Sources & Behaviors

Based on these considerations, the availability of videos for significant investors could represent a value that could be regulated by the simplified relationship:

$$Vd = IBvf + GBvf + IPvf + GPvf$$

$$IBvf = k_{gb} \cdot IBv \left[cIB_0 - \frac{(1 - cIB_0) \cdot \log(sa + 10)}{\mathrm{cIB}_a \cdot (\mathrm{sa} + 1)} \right]$$

$$GBvf = k_{gb} \cdot GBv \left[cGB_0 - \frac{(1 - cGB_0) \cdot \log(sa + 10)}{\mathrm{cGB}_a \cdot (\mathrm{sa} + 1)} \right]$$

$$IPvf = k_{gb} \cdot IPv \left[cIP_0 - \frac{(1 - cIP_0) \cdot \log(sa + 10)}{\mathrm{cIP}_a \cdot (\mathrm{sa} + 1)} \right]$$

$$IPvf = k_{gb} \cdot GPv \left[cGP_0 - \frac{(1 - cIP_0) \cdot \log(sa + 10)}{\mathrm{cGP}_a \cdot (\mathrm{sa} + 1)} \right]$$

IBv Value of acquiring knowledge on the Individual Behavior for supporting future Financial Deals with the subject

GBv Value of acquiring knowledge on the Generalized Behavior for supporting future Financial Deals with potential customers

IPv Value of acquiring knowledge on the Individual Profiling for targeting various products/services on the subject

GPv Generalized Profiling for identification of potential customers for targeting various products/service

IPvf Final Asymptotic value for Individual Behavior

GBvf Final Asymptotic value for Generalized Behavior

IPvf Final Asymptotic value for Individual Profiling

GPvf Final Asymptotic value for Generalized Behavior

cXX_0 Initial value in terms of percentage respect XX (i.e., IB, GB, IP, GP)

cXX_a speed in reaching the final asymptotic value

k_i Corrective factor for the i-th specific element

sa saturation of data measuring the coverage of the available data set to respect that is supposed to obtain the total value just when reaching +oo

As already mentioned, we suggest using zero for k_{ip} in order to reduce risks related to privacy; so the residual values could be estimated as follows:

$$GBv = \sum_{j=1}^{m_{gb}} NGB_j \cdot efGB_j \cdot aGBBV_j \cdot pGBBM_j$$

$$GPv = \sum_{j=1}^{m_{gp}} NGP_j \cdot efGP_j \cdot aGPBV_j \cdot pGPBM_j$$

$$IBv = NI \sum_{j=1}^{m_{gp}} efIP \cdot aIPBV_j$$

NGB_j j-th quantity of potential customers for j-th business related to Generalized Behavior

NGP_j j-th quantity of potential customers for j-th business related to Generalized Profiling

NI number of internal recorder subjects

$efGB_j$ efficiency in percentage for improving the j-th business related to Generalized Behavior

$efGP_j$ efficiency in percentage for improving the j-th business related to Generalized Profiling

$efIP_j$ efficiency in percentage for improving the j-th business related to the Individual Profiling

$aGBBV_j$ Average Value of the j-th business related to a person in terms of the Generalized Behavior

$aGPBV_j$ Average Value of the j-th business related to a person in terms of the Generalized Profiling

$aIPBV_j$ Average Value of the j-th business related to the subject in terms of the Generalized Profiling

$pGBBM_j$ Penetration on the market for the j-th business for the Generalized Behavior

$pGPBM_j$ Penetration on the market for the j-th business for the Generalized Profiling

The quantification of NXX, aXXBV, and pXXBM could be attributed to specific business sectors; for instance, the beauty product market, pharmaceutical or entertainment; while the estimation efXX is related to the capability to extract valuable information from the different data sources (i.e., video, audio, transaction and deal records, customer profile) that relies on the capabilities of the proposed architecture. To provide an example in case, we could achieve a 60% efficiency in identifying interest in the high-end jewelry market of the investors recorded in this case and keeping to zero k_{ip}; we can estimate one of these elements such as jewelry in the USA as jw (US Census, 2016; Jeffay, 2017).

For instance, the $aGBBM_{jw}$ could be evaluated equivalent to 7285 USD/year per person (over 70k USD annual incomes) and correspond to a market of around 58 million people corresponding to 41bUSD; while moving up to high and medium jewelry, the estimate of the $aGBBM_{jw} \cdot pGBBM_{jw}$ turns to correspond around to USD 6 Bn in the USA. This means that if by data we can improve 1% sales on the wealthy customers, we have a value of around 60 MUSD as improvement of the existing market, while the potential to the growth of 1‰ of jewelry expenditure could extend the market of 41m USD. Considering the reliability of just 30% that could be pretty conservative, these data could have a total impact on this single business of around 30m USD/year. So, it turns evident that the value should be considered the value of corrective factors properly. In the specific case, the limited data is partially compensated by fitting high investors as potential customers for these products. Therefore the value proposed to represent, corrected by kGB and kGP, the asymptotic value for these data as assets; experimentation in using this information and measuring business revenues could provide not only a more effective estimation on their value but also to create models that could correspond to the estimation of their evolution based on the extension of the data set.

5. Conclusions

The proposed investigation represents the first step forward to identify and develop new methodologies for evaluating the potential of Data in different areas to transform them into tangible financial assets for Companies operating within particular sectors. It is evident that if the data number is small, there is an evident limited statistical value. Therefore, introducing these concepts and evaluations could further extend the data set and fuse different types and sources to increase the corresponding value. In this case, the analysis suggests that it could be possible to reuse experiences applied in other fields to create deductive capabilities even from relatively small quantities of data.

References

Abdelhedi, S., Wali, A., & Alimi, A. M. (2016). Fuzzy logic-based human activity recognition in video surveillance applications. In Proceedings of the Second International Afro-European Conference for Industrial Advancement AECIA 2015 (pp. 227-235). Springer, Cham.

Abdulsalam, W. H., Alhamdani, R. S., & Abdullah, M. N. (2019). Facial emotion recognition from videos using deep convolutional neural networks. International Journal of Machine Learning and Computing, 9(1), 14-19.

Achara, J.P., Acs, G. and Castelluccia, C., 2015, October. On the unicity of smartphone applications. In Proceedings of the 14th ACM Workshop on Privacy in the Electronic Society (pp. 27-36).

Aho, B., & Duffield, R. (2020). Beyond surveillance capitalism: Privacy, regulation and big data in Europe and China. Economy and Society, 49(2), 187-212.

Aridor, G., Che, Y. K., & Salz, T. (2020). The economic consequences of data privacy regulation: Empirical evidence from GDPR. NBER working paper, (w26900).

Bhardway J., Long J. (2021) "What are the different kinds of data profiling?", Proc. of Open Data Science Conference, September 15-16

Bloor R. (2020) "How Much Is Data Worth? The Value of Your Personal Data", Permission, April 8

Blumenstock, J., Cadamuro, G. and On, R., 2015. Predicting poverty and wealth from mobile phone metadata. Science, 350(6264), pp.1073-1076.

Broekens, J., Jonker, C. M., & Meyer, J. J. C. (2010). Affective negotiation support systems. Journal of Ambient Intelligence and Smart Environments, 2(2), 121-144.

Bruzzone, A. G., Massei, M., Sinelshchikov, K., Fadda, P., Fancello, G., Fabbrini, G. & Gotelli, M. (2019). Extended reality, intelligent agents, and simulation to improve efficiency, safety, and security in harbors and port plants. In 21st International Conference on Harbor, Maritime and Multimodal Logistics Modeling and Simulation, HMS 2019 (pp. 88-91). DIME University of Genoa.

Bruzzone, A. G., Di Matteo, R., & Sinelshchikov, K. (2018) Strategic Engineering & Innovative Modeling Paradigms. In Workshop on Applied Modelling & Simulation, Praha, CZ

Bruzzone, A.G., Massei, M., Sinelshchikov, K. & Di Matteo, R. (2018). Population behavior, social networks, transportations, infrastructures, industrial and urban simulation. Proceedings of 30th European Modeling and Simulation Symposium, EMSS 2018, Budapest, Hungary. pp. 401-404

Bruzzone, A., Massei, M., Longo, F., Poggi, S., Agresta, M., Bartolucci, C. & Nicoletti, L. (2014). Human behavior simulation for complex scenarios based on intelligent agents. Proceedings of ANSS2014, Spring Simulation Multi-Conference (SpringSim'14) April 13 – 16, Tampa, FL; USA

Bughin, J., Chui, M., & Manyika, J. (2010). Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. McKinsey Quarterly, 56(1), 75-86.

Cavanillas, J. M., Curry, E., & Wahlster, W. (2016). The big-data value opportunity. In New horizons for a data-driven economy (pp. 3-11). Springer, Cham.

Chakraborty, A., Konar, A., Chakraborty, U. K., & Chatterjee, A. (2009). Emotion recognition from facial expressions and their control using fuzzy logic. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 39(4), 726-743.

Chittaranjan, G., Blom, J. and Gatica-Perez, D., 2011, June. Who is who with big-five: Analyzing and classifying personality traits with smartphones. In Wearable Computers (ISWC), 2011 15th Annual International Symposium on (pp. 29-36). IEEE.

Crowley-Sweet Davin (2021) "Proving Data's Strategic Value To Get You Board-Level Buy-In," DataIQ

De Silva, L. C., Miyasato, T., & Nakatsu, R. (1997, September). Facial emotion recognition using multi-modal information. In Proceedings of ICICS, 1997 International Conference on Information, Communications and Signal Processing. Theme: Trends in Information Systems Engineering and Wireless Multimedia Communications (Cat. (Vol. 1, pp. 397-401). IEEE.

Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. Computer, 51(8), 56-59.

Jeffay J. (2017) "Who Spends on Jewelry: Income Levels Are Key," Index Online, February 8

Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W. and Torralba, A., 2016. Eye-tracking for everyone. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2176-2184)

Kaya, M. F., & Schoop, M. (2019, June). Application of data mining methods for pattern recognition in negotiation support systems. In International conference on group decision and negotiation (pp. 223-237). Springer, Cham.

Laney D.B. (2021) Data Valuation Paves The Road To The Future For Highways England", Forbes, February 1

Leonelli, S. (2019). Data—from objects to assets, Nature, October, 574, pp. 317-320

Maier, J., Schlechte, D., Fernandes, M., & Theissler, A. (2020). A deep learning approach to prepare participants for negotiations by recognizing emotions with voice analysis. In Proceedings of the 20th International Conference on Group Decision and Negotiation, Toronto, Canada.

Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. International journal of human-computer studies, 65(4), 329-347.

Mavriki P, Karyda M. (2018) "Profiling with big data: Identifying privacy implications for individuals, groups, and society," Proc. of MCIS, Corfu, Greece

Nemitz, P. (2018). Constitutional democracy and technology in the age of artificial intelligence. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2133), 20180089.

Nicola, F. G., & Pollicino, O. (2020). The Balkanization of Data Privacy Regulation. W. Va. L. Rev., 123, 61.

Pantic, M., & Rothkrantz, L. (2004, June). Case-based reasoning for user-profiled recognition of emotions from face images. In 2004 IEEE International Conference on Multimedia and Expo (ICME)(IEEE Cat. No. 04TH8763) (Vol. 1, pp. 391-394). IEEE.

Pennebaker J.W., M.R. Mehl, K. Niederhoffer (2003) Psychological aspects of natural language use: Our words, our selves, Annual Review of Psychology (54) (2003), pp. 547-577

Ploug, T., & Holm, S. (2021). The Right to Contest AI Profiling Based on Social Media Data. The American Journal of Bioethics, 21(7), 21-23.

Poyraz, O. I., Canan, M., McShane, M., Pinto, C. A., & Cotter, T. S. (2020). Cyber assets at risk: monetary impact of US personally identifiable information mega data breaches. The Geneva Papers on Risk and Insurance-Issues and Practice, 45(4), 616-638.

Rangel, F., & Rosso, P. (2016). On the impact of emotions on author profiling. Information processing & management, 52(1), 73-92.

Rossi, F., Fasel, I. and Sanfey, A.G., 2011. Inscrutable games? Facial expressions predict economic behavior. BMC Neuroscience, 12(1), p.P281.

Saha, A. K. (2017). Review of Design of Speech Recognition and Text Analytics based Digital Banking Customer Interface and Future Directions of Technology Adoption. Computer Science, December, arXiv preprint arXiv:1712.04640.

Schler, J., Koppel, M., Argamon, S., & Pennebaker, J. W. (2006, March). Effects of age and gender on blogging. In AAAI spring symposium: Computational approaches to analyzing weblogs (Vol. 6, pp. 199-205).

Sengupta S., Rusli E.M. (2012) "Personal Data's Value? Facebook Is Set to Find Out", The New York Times, January 31

Singh, V.K., Freeman, L., Lepri, B. and Pentland, A.P., 2013. Classifying spending behavior using socio-mobile data.

Sjöberg, M., Chen, H. H., Floréen, P., Koskela, M., Kuikkaniemi, K., Lehtiniemi, T., & Peltonen, J. (2016, September). Digital me: Controlling and making sense of my digital footprint. In International Workshop on Symbiotic Interaction (pp. 155-167). Springer, Cham

Sottilare, R. A., & Proctor, M. (2012). Passively classifying student mood and performance within intelligent tutors. Journal of Educational Technology & Society, 15(2), 101-114.

US Census Bureaus (2016) "Person Income in 2015 - 15 Years and Over - All Races", Technical Report US Census Bureau. 2016, Suitland, MD

Williams N. (2017)"What exactly is 'profiling' under the GDPR," Data & Marketing Association, June 20

Wimmer, K. (2018). Free expression and EU privacy regulation: Can the GDPR reach US publishers. Syracuse L. Rev., 68, 547.

Wingenbach T. S., C. Ashwin, and M. Brosnan, (2016) "Correction: Validation of the Amsterdam dynamic facial expression set–bath intensity variations (ADFES-BIV): A set of videos expressing low, intermediate, and high-intensity emotions," PloS One, vol. 11, no. 12, p. e0168891

Wu, H.Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F. and Freeman, W., 2012. Eulerian