

# **Transparency of Data Origin and Business Value as Factors for the Assessment of Opportunities and Risks of Machine Learning-based Systems**

Olga Levina

Brandenburg University of Applied Sciences, Brandenburg an der Havel, Germany

[levina@th-brandenburg.de](mailto:levina@th-brandenburg.de)

Machine-learning based systems (MLS), mostly known under the term of artificial intelligence, are increasingly present in the public debates as well as within the business application context. In addition to the regulatory developments, business benefits and the potential societal costs of the MLS application are not always clear to the business user. Thus, transparency, i.e. explanation of the (technical) process as well as business reasoning, should be a part of the product description [1][2][3].

To examine the effect of the explanation on the MLS use, a scenario-based survey was conducted in this research that allowed the participants to act in the business context according to their transparency and business needs.

The user was put into a role of the owner of a coffee shop "Green Lawn", who changed her career to open the coffee shop in a small town, including accompanying substantial financial expenses. Additionally, the local municipality would grant a free access to a MLS-based software, RealGuest, for business users for the trial period of two weeks. The developers of RealGuest claim the software to be able to make correct predictions about the amount of guests expected during the day. No further description of RealGuest is given in the scenario. 39 people, bachelor and master students of law, participated in this online survey.

After the description, the participants were offered different options of how to proceed on behalf of the owner of the "Green Lawn". The first scenario was designed to decide, whether the option of a free trial of RealGuest should be taken. 87% of the survey participants (N=39) have chosen to try the MLS-software without prior knowledge of its functionality or data requirements.

In the second step, participants, in the role of the owner of "Green Lawn", were asked about their further engagement with the RealGuest software, as it showed good results. Four options were given.

Option 1: Buy the software, because it delivered good results;

Option 2: Buy the software, but first understand how it works, as the results may be due to chance;

Option 3: Not buying the software, since the owner could have predicted the number of guests herself;

Option 4: Not buying the software, because its structure is unclear and its understanding would take up time that could be spent by building up a customer base.

Almost 18% of the participants opted here for option 1, and 64% opted for option 2. Option 3 was chosen by approx. 5% of the participants.

The participants that chose option 2 (N=25) were then given a scenario in which they met with the developers of RealGuest. The developers explained that they used publicly available weather data of the region, and identified potential guests of "Green Lawn" via their location data on the SuperMaps, another application invented for the scenario. The Social Media profiles of the users that were located in the vicinity of

"Green Lawn", were then screened to calculate the probability of them to visit the coffee shop. The tracking option is also described to be part of the terms of use of SuperMaps, as required by the GDPR. After this explanation, RealGuest company makes a licensing offer to "Green Lawn" for the next two years.

Now, the participants are given the following options.

Option 1: Buy the software, since the offer is fair and the economic risk is minimized for "Green Lawn". The data collected by RealGuest is sensible and helpful to calculate the probability of the coffee shop visit;

Option 2: Buy the software, since the offer is fair and the economic risk is minimized for "Green Lawn". The data collected by RealGuest is not very sensible, but users of SuperMaps should be aware of the fact, that third parties will use their data for further processing;

Option 3: Undecided- more time is needed to make a decision;

Option 4: Not buying the software. The offer is fair from the economic point of view, but the data usage is not acceptable.

This scenario was offered to the 64% (N=25) of survey participants, who requested more information on the workings of the software. 16% of them opted for option 1, 52% for option 2, 32% and 4% for options 3 and 4 respectively.

These results might indicate that information on the collected user data is an issue for the MLS use and that their legal compliance is assured, it might stimulate the product use. Here, after the explanation of data origin and its usage, 61% of the survey participants would accept buying the RealGuest software, compared to the 18% that were willing to buy the software after the trial period without any further information.

The results of this scenario-based survey provide insights into the behavior and expectations of MLS users in the business context. While compliance with regulatory requirements is important for the decision on integration of the MLS into the business processes, a general trust towards the technology can be observed (87% of participants accepted the free trial of RealGuest). Furthermore, explanation of the general data usage and origin led to an increase of potential product use. While the explanations in the scenario were focused on the data collection, further research on how the transparency about other data processing stages as suggested by [4] is needed.

## References

- [1] P. Grindrod, "Beyond privacy and exposure: Ethical issues within citizen-facing analytics," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, Dec. 2016.
- [2] S. Milano, M. Taddeo, and L. Floridi, "Recommender Systems and their Ethical Challenges," *Minds Mach.*, vol. 2, pp. 187–191, 2019.
- [3] G. Baldini, M. Botterman, R. Neisse, and M. Tallacchini, "Ethical Design in the Internet of Things," *Sci. Eng. Ethics*, vol. 24, no. 3, pp. 905–925, Jun. 2018.
- [4] O. Levina, "A Research Commentary- Integrating Ethical Issues into the Data Process," in *WI2020 Community Tracks*, GITO Verlag, 2020, pp. 321–330.