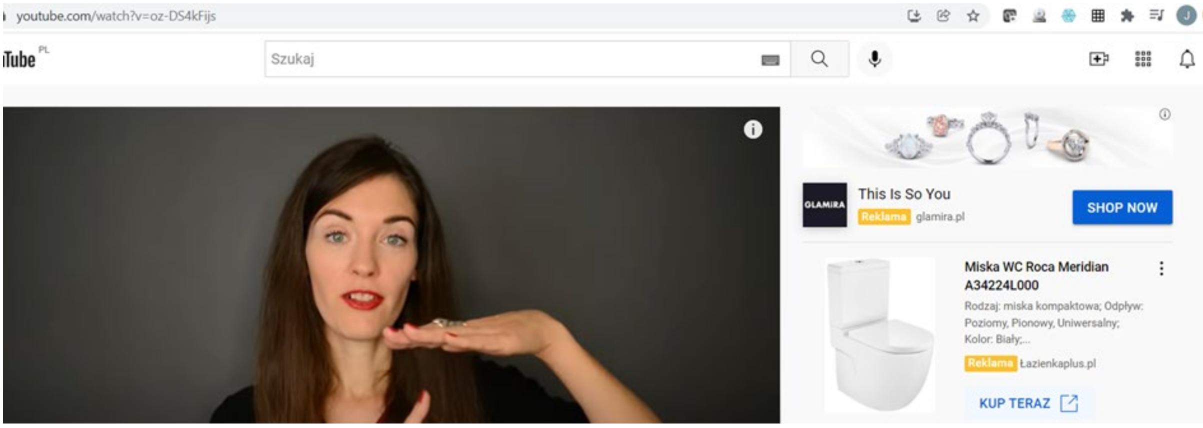
Pushy Toilet sellers and other problems of the recommender systems

If the reader, like me, faced the challenge of buying a toilet compact at some point in his life, he knows what a review of several models in an online store means for the algorithms that offer us the advertising: “Here he is: a real connoisseur of toilet compacts!”. “I see you are browsing through your friend’s feed on Facebook, this must be a good time to get you interested in this beautiful white model.”, “Are we watching some video? How about seeing some more toilets?”



You might even be interested if not for the fact that you’ve already bought one a week ago and don’t really have a space for these lovely new models. This is an example of one of the very common biases we meet when we are using the recommender systems: the **Inductive Bias.** Inductive bias denotes the assumptions made by the model to better learn the target function and to generalize beyond training data. In this particular case, the recommender system is built with an implicit assumption that the person that purchased a certain item is more likely to purchase it in the future. The general assumption doesn’t take into account the one-off nature of the purchase.

Types of biases in the recommender systems

As noted in a comprehensive review of a literature on bias in recommended systems by Chen and colleagues \cite{chen2020bias}, recent years have seen a surge of research effort on recommendation biases, however, the term is used inconsistently across papers and is used to describe a series of different phenomena.

One type of bias is the one in **explicit feedback data** - the data that store how certain users rated certain items, for example, the databases with restaurants and numbers of stars or pluses user's given each of them. There are two problems with such datasets. First is selection bias - users are free to choose which items to rate, so that the observed ratings are not a representative sample of all ratings. The thing is we usually want to express an opinion either when something is exraordinary pleasing or on the contrary, extraordinary dissapointing. This means that the restaurants that are simply ok, not very good or bad receive less ratings. This means also the ratings given will tend to be concentrated on the both extremes of the scale. Other reason, why the explicit feedback data might not always show the true picture is a conformity bias. Let's take an example of a movie that is very popular both among your group of friends and film critiques. If you are prone to the conformity bias (and almost all of us are), you could feel hesitant putting a low rating, even if you didn't enjoy it.

Unfairness

Another widely disputed topic in the recommender system's world is the issue of fairness and discrimination. Although it seems odd to suspect a machine of some hidden prejudices this is a serious obstacle to making recommender systems more entrenched within our society. In particular, based on attributes like race, gender, age, education level, or wealth, different user groups are usually **unequally represented in data**. When such data is used as input to train a model, the models will better serve the over-represented groups, reinforce them in the ranked results, and potentially result in **systematic discrimination**. For example, in the context of job recommendation, previous work found that, compared to men, women saw fewer ads about more prestigious jobs and career coaching services.

Self-Induced Bias

The subject of author's of this article research is a Self-Induced Bias (SIB). The recommender systems are all based on the same mechanism - the decisions from the past are used to train the system and if used, the system will strengthen this type of decision. The systems don't need to know anything about who are you, what are your interests and what kind of things it is recommending, the same algorithm can be used to propose a washing machine, holiday destinations, and art pieces. This is both the recommender system's strength and weakness. Now imagine that you are opening a brokerage account and start investing your money in stocks. First, you pick stock A, which falls 10% after announcing some bad earning prognosis. Then you pick stock B, which collapses two months later due to fraud. Lastly, you invest the money you have left in company C, which also doesn't work out well. If the recommender system got your track record without any context it would learn that you simply prefer the stocks that have a high chance to drop, and, if it was a collaborative filtering type of system, it would match you with other users with similar bad luck. This example, although exaggerated shows, how the user can provide the recommender system with input, based on which the system will actually cause more harm to the user with its propositions. Not the best advisor, right?