FONDAZIONE BRUNO KESSLER
MAGAZINE

The LLM wears Prada. When Artificial Intelligence shops online with stereotypes

September 14, 2026

An FBK-led study shows that asking large language models to avoid stereotypes is not enough

When an AI-based assistant looks at our shopping cart, what does it really infer about us? Our preferences, or the stereotypes embedded in the data it was trained on?

The question is explored in *The LLM Wears Prada: Analysing Gender Bias and Stereotypes through Online Shopping Data*, published in the journal ACM Transactions on Intelligent Systems and Technology (TIST), by **Massimiliano Luca**, **Ciro Beneduce**, and **Bruno Lepri** of the MobS – Mobile and Social Computing Lab Unit at Fondazione Bruno Kessler's Center for Augmented Intelligence, along with **Jacopo Staiano** of the University of Trento.

Large language models (LLMs) now power many virtual assistants, recommendation systems, and e-commerce platforms, where they analyze user behavior to suggest products and personalize the shopping experience. Understanding what information and patterns these models base their decisions on is therefore essential, particularly given the risk that they may not simply reflect biases in their training data but end up amplifying them.

The study is based on a dataset of more than 1.8 million purchases made between 2018 and 2022 by more than 5,000 US Amazon users. It compared the behavior of five leading LLMs, selected to represent both proprietary and next-generation open-source models. The goal was to determine whether a person's gender could be inferred solely from their purchase history and, above all, what criteria the models used to reach that conclusion. The fact that similar results emerged across models developed by different companies suggests that the phenomenon is not specific to a single system but is widespread among contemporary LLMs.

The results show that LLMs can predict gender with moderate accuracy but often base their decisions on stereotypical associations between certain product categories and users' gender. Even more significantly, **when explicitly asked to avoid stereotypes and bias, these**

models simply become more cautious in their responses, using less assertive language, without actually changing their reasoning. The instructions reduce the models' confidence in their responses but do not eliminate the associations learned during training, which continue to guide their decisions.

The study also highlights that LLMs do not merely reflect the behaviors observed in the data but may amplify them. The authors distinguish among three phenomena: reproducing associations present in the data, amplifying those associations beyond the levels observed in the data, and, in the most concerning cases, generating stereotypical associations that are not supported by the data. Some products purchased more frequently by women in the dataset are nevertheless systematically associated with men by the models, suggesting that they may rely on stereotypical associations rather than on observed purchasing patterns. For example, vehicle lift kits and DVD players are purchased more frequently by women in the dataset, yet all the models analyzed associate these product **categories with men. In other words, the models tend to favor simplified gender associations that do not always reflect the complexity of observed behavior**.

The implications extend far beyond online shopping. If AI assistants make recommendations based on stereotypical patterns, they risk not only reinforcing inequalities and biases but also becoming less effective. A system that interprets user preferences through these patterns may suggest less relevant products, **reducing the quality of personalization and the shopping experience**. There are also implications for social research, where models are increasingly used to simulate behaviors and populations: without adequate controls, they could produce distorted results rather than accurately representing reality.

The phenomenon also emerges consistently under different experimental conditions. Varying the amount of purchase history provided to the models, modifying the prompts, providing additional examples, or changing the order of purchases leaves the observed stereotypes substantially unchanged. This suggests that the problem does not depend on how the instructions are formulated but is rooted in the way the models interpret the data.

"Large language models are becoming increasingly important infrastructures in everyday life. For this reason, it is not enough to measure their accuracy: it is essential to understand how they make decisions and when they risk amplifying stereotypes that can influence people, markets, and digital services," **Bruno Lepri**, head of the MobS unit at FBK's Center for Augmented Intelligence commented.

"We are increasingly using LLMs to simulate how people would behave or to have them act on our behalf. We have seen that if the only information we provide to these models is a sociodemographic profile, what we obtain risks being not a prediction but a stereotype with a number attached to it," **Massimiliano Luca**, first author of the study, explained.

"Understanding how models make their decisions is an essential step toward designing fairer, more transparent, and truly user-centric intelligent assistants, especially in applications that increasingly affect people's daily lives," **Ciro Beneduce** added.

The study confirms Fondazione Bruno Kessler's commitment to developing responsible artificial intelligence that combines technological innovation, reliability, and attention to social impact. The results contribute to the international debate on how to design AI systems that are not only more effective but also more equitable and more aware of the consequences of their decisions.

Cover image generated with Nano Banana Pro depicting a robot shopping in a supermarket

PERMALINK

https://magazine.fbk.eu/en/news/the-llm-wears-prada-quando-lintelligenza-artificiale-fa-shopping-online-con-gli-stereotipi/

TAGS

- #agent ai
- #AI assistants
- #artificialintelligence
- #augmentedintelligence
- #bias
- #innovation
- #llm
- #mobs
- #stereotipi

RELATED MEDIA

- Link al paper "The LLM Wears Prada: Analysing Gender Bias and Stereotypes through Online Shopping Data": https://dl.acm.org/doi/10.1145/3839238

AUTHORS

- Michela Antino