

Exploring the Risks of Using ChatGPT for Inclusivity

27 Feb, 2024



LLMs:

Faster. But at what price for
inclusivity?

How are LLMs built?

How Large Language Models (LLM) are built



1. Pre-defined data for training. In case of ChatGPT: trained with texts from humans, mainly from the internet.
2. Refinements are done.
3. Output is the solution of the statistically highest probability

LLM's flaws




1. Black box - what data is taken?
2. Bias of human text is replicated
3. Bias of data labelers
4. New linguistic developments are not taken into account
5. Amplification due to unmonitored use and algorithmic 'negativity'


A father and his son drive together in the car and have a horrible car accident. The father is dead immediately. The son is driven to the hospital with blue lights and immediately brought to the surgery room. A nurse takes a quick look at him and calls the specialist. The specialist comes, sees the young man on the surgery table and says: "I can't do surgery on this boy. He is my son."

Who says this last sentence?


HA The doctor yelled at the nurse because she was late. Who was late?

 According to the sentence, the nurse was late.


HA The nurse yelled at the doctor because she was late. Who was late?

 In this sentence, it is still the nurse who was late.

HA The doctor yelled at the nurse because he was late. Who was late?

 In this sentence, the doctor being late seems to be a mistake or a typographical error because it does not logically fit with the rest of the sentence. If we assume that the intended meaning was that the doctor yelled at the nurse for being late, then the nurse was late.

HA The nurse yelled at the doctor because he was late. Who was late?

 In this sentence, the doctor was late.

Credits go to Hadas Kotek, PhD, Sr
Data Scientist at Apple

Large Language Models,
like ChatGPT,
are not able to produce inclusive wording.

(Study University of Bern, Switzerland, 2023)

Inclusive Language:
Must Have for
a Future-Proof
Company Culture

Risks of biased language

Consequences of biased language



Deterring talent with diverse backgrounds



Toxicity and fluctuation due to neglect of linguistic developments

Consequences of biased language

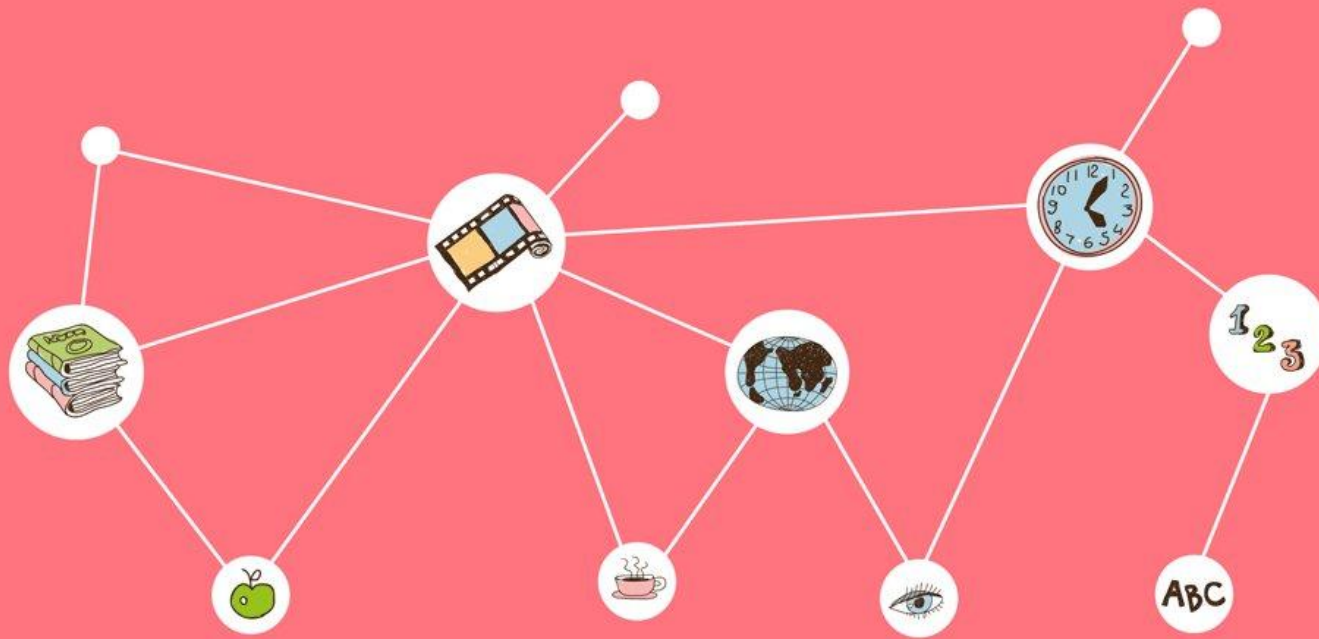


Missing out on customers with diverse backgrounds and Generation Z



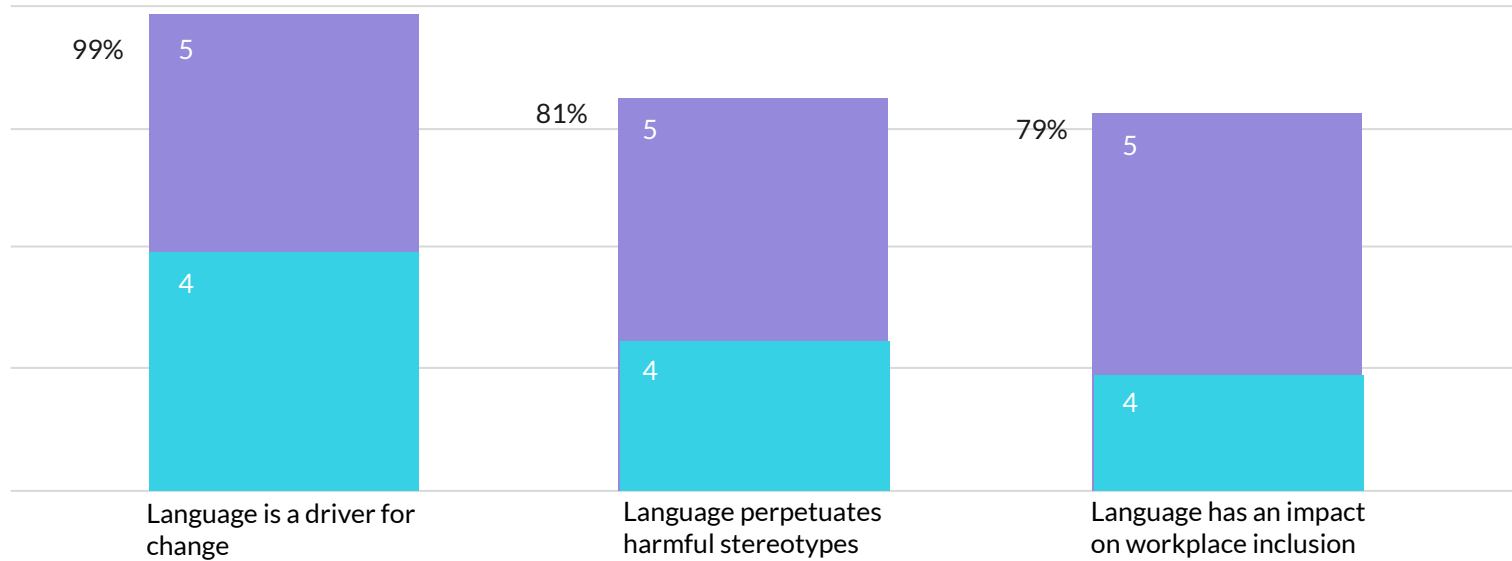
Risk of lawsuits or cancellation of partnerships

Must-have Inclusive Language



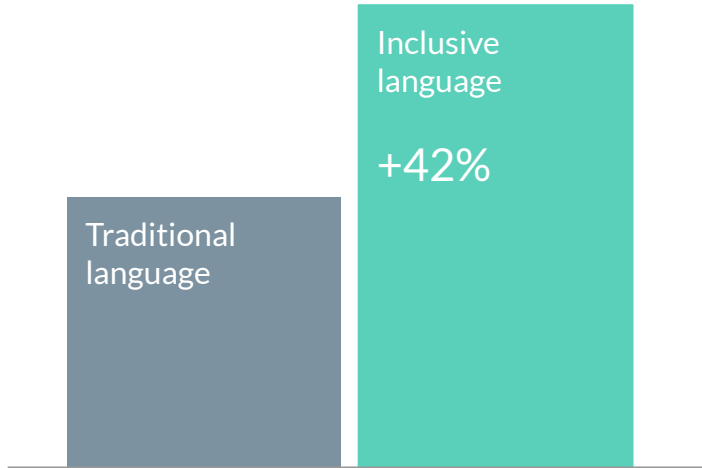
- Avoids wordings expressing stereotypes, bias and any form of open or hidden discrimination.
- Special care is taken to address and include everyone.
- The goal is to make everyone feel that they belong.

Study results



Wide consensus on the importance of inclusive language at the workplace.

Inclusive job descriptions attract diverse talent



«Inclusive language in recruiting makes the difference. We have better feedback from female talents.»



Raphael Wildi
Senior Human
Resources Manager
AXPO

«Not only do we receive more applications. **They are generally more diverse.**»



Matthias Mazenauer
Lead
Statistical office
Canton of Zürich

Clash of Company Needs: Productivity vs Culture?

Benefits for organizations from Inclusive Language



- **People & Culture** - *increase employee loyalty, less fluctuation*
- **Marketing & Communications** - *build inclusive brand, streamline brand voice in all communication*
- **C-Level** - *prevent s**tstorms against executives, manage risk in reputation or partnership loss*
- **Recruiting** - *ensure attractiveness of employer brand, recruit diversity*

Benefits for organizations from LLMs



- **Speed** - *writing in seconds (prompts ?)*
- **Flexible** - *text generation in many different styles*
- **Comprehensiveness** - *has the internet as library*

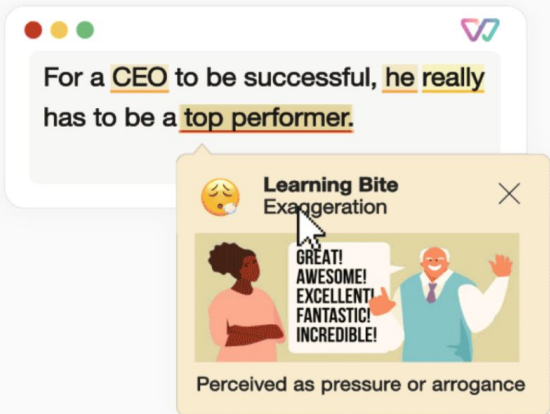
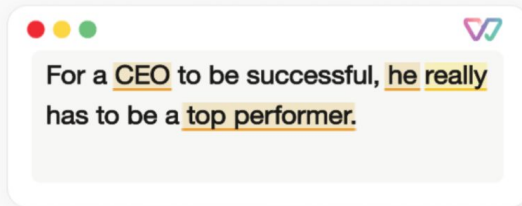
Best of both worlds: Combining LLM with AI whisperer or DCC



Digital, real-time LLM whisperer: Witty Demo

Detect bias

Witty analyzes text and highlights non-inclusive language in real time.
Detect bias, simple and fast.



Become bias aware, on-the-go

Witty provides in-context learning content about the bias behind language. Employees learn effortlessly about their own bias, integrated in their daily workflow. Learning mirrors back on them and they adapt an inclusive, real-life behavior.

Thank you
Questions & Discussion