

# Artificial Intelligence and Discrimination

**Authors:** Kimon Kieslich, Christopher Starke, Pero Došenović, Birte Keller, & Frank Marcinkowski

## Key Findings

**Discrimination through Artificial Intelligence (AI) is generally only perceived as a moderate risk by the population. However, when it comes to negative consequences for the economy, the use of AI is viewed rather critically. Many citizens would like to see stronger regulation of AI.**

## Background

The fact that discrimination is a central social problem was made clear not least by the Black Lives Matter demonstrations taking place worldwide. But it is not only people of color who are affected by discrimination - every day, people are discriminated against on the basis of their gender, their faith, their sexuality or their family background. Recently, systems with artificial intelligence have also repeatedly come into focus in connection with discrimination. In the past, there have been examples of discriminatory AI applications in personnel recruitment, in police work or in connection with the utilization of chatbots (Beck et al., 2019).

As can be seen on the official [MeMo:KI] Dashboard, the German population is generally supportive of the use of AI, at least in some areas of application. However, it is unclear whether German citizens are aware of the risk potential of AI technologies and in which specific areas of application citizens even suspect discrimination potential through AI. Furthermore, the

question arises as to how discrimination can be prevented. In this context, the EU Commission (2020) has drawn up guidelines for trustworthy AI, which propose various countermeasures to combat discrimination by AI. For a political legitimization of these countermeasures, it is therefore also relevant to record to what extent citizens consider them to be effective and which ones should be implemented most urgently.

In the latest survey of the [MeMo:KI], 1,022 German citizens were interviewed on the topic of "AI and discrimination". The central results are presented in this factsheet. Since a basic understanding of artificial intelligence is a prerequisite for an evaluation of the questions, only citizens who stated that they know what the term means or could possibly explain it were taken into account in the following analyses. In this survey, this applied to 915 people (89.5% of respondents).

## Methodology

<b>Method</b>	Online Survey
<b>Executing Institute:</b>	forsa Politik & Sozialforschung GmbH
<b>Population:</b>	German population over 18 years of age who use the Internet at least occasionally
<b>Sample:</b>	Weighted random sample (N=1,022)
<b>Weighting Criteria:</b>	Age, gender and region (federal state)

Survey Period: 2020, July, 27-31

Further Information: [Detailed Methodology Overview](#) for the MeMo:KI project [in German language].

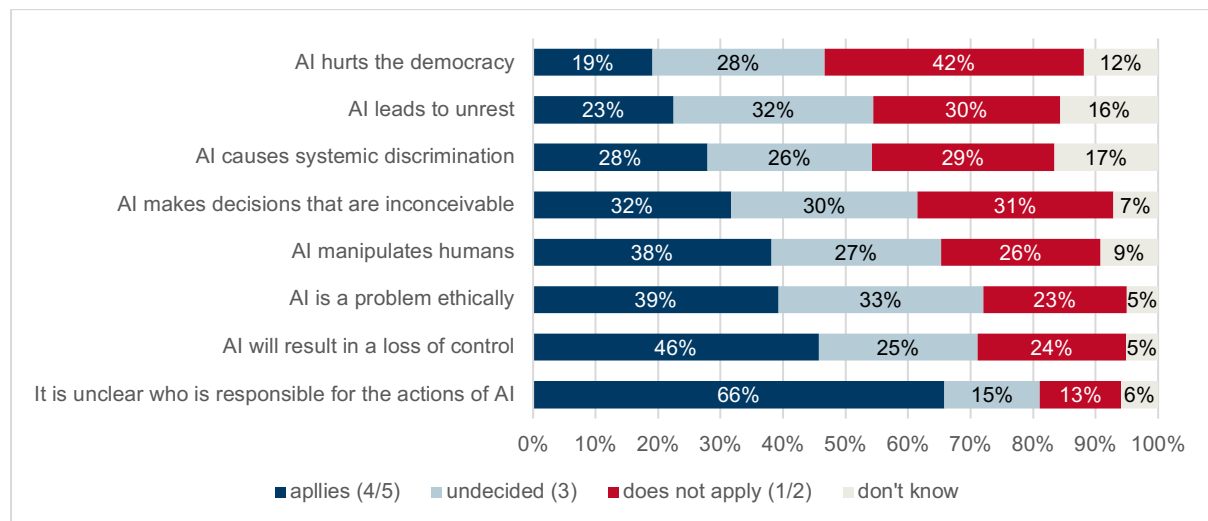
## Respondents see only medium risks regarding discrimination through AI

Discrimination is one of many possible risks that AI technology can bring with it. We asked German citizens how strongly they rate various problems in the use of AI for society. They were able to rate their opinion on a five-point scale, with (1) representing "do not apply" and (5) representing "fully apply". For the evaluation, the gradings (1) and (2) were combined to represent "no risk perception" and (4) and (5) were combined to represent "risk perception".

The data show that discrimination (response options "systematic disadvantage") is seen as a great or very great risk by only around 28% of

respondents. By contrast, the proportion of those who (tend to) see no risk of discrimination in the use of AI slightly outweighs this at 29%. Strongly perceived risks, on the other hand, are unclear accountability in the use of AI technologies (66% agreement) or a possible loss of control (46% agreement). However, the fact that AI can cause injustice (23% approval) or even harm democracy (19% approval) is assessed as a low risk. Overall, it can be seen - especially in comparison to the other risk potentials - that discrimination through AI is rated by the respondents as a minor problem.

Figure 1: Risk perception towards AI technology



Annotation: N=913

Question: Now we are also interested in your personal opinion on the influence of artificial intelligence on society. Please rate to what extent the following statements apply or do not apply to AI. (1=does not apply to; 5=fully applies)

## Concern about discrimination on economic issues

Discrimination takes place in different areas of society. In 2017, the Federal Anti-Discrimination Agency published a report on the status quo of discrimination in Germany, listing problematic areas, among others - AI systems are  
A project of:

also used in many of these areas today (Anti-diskriminierungsstelle des Bundes, 2017). In addition, the European Council commissioned a legal assessment on the potential for discrimination through the use of AI (Borgesius, 2018).

Funded by:

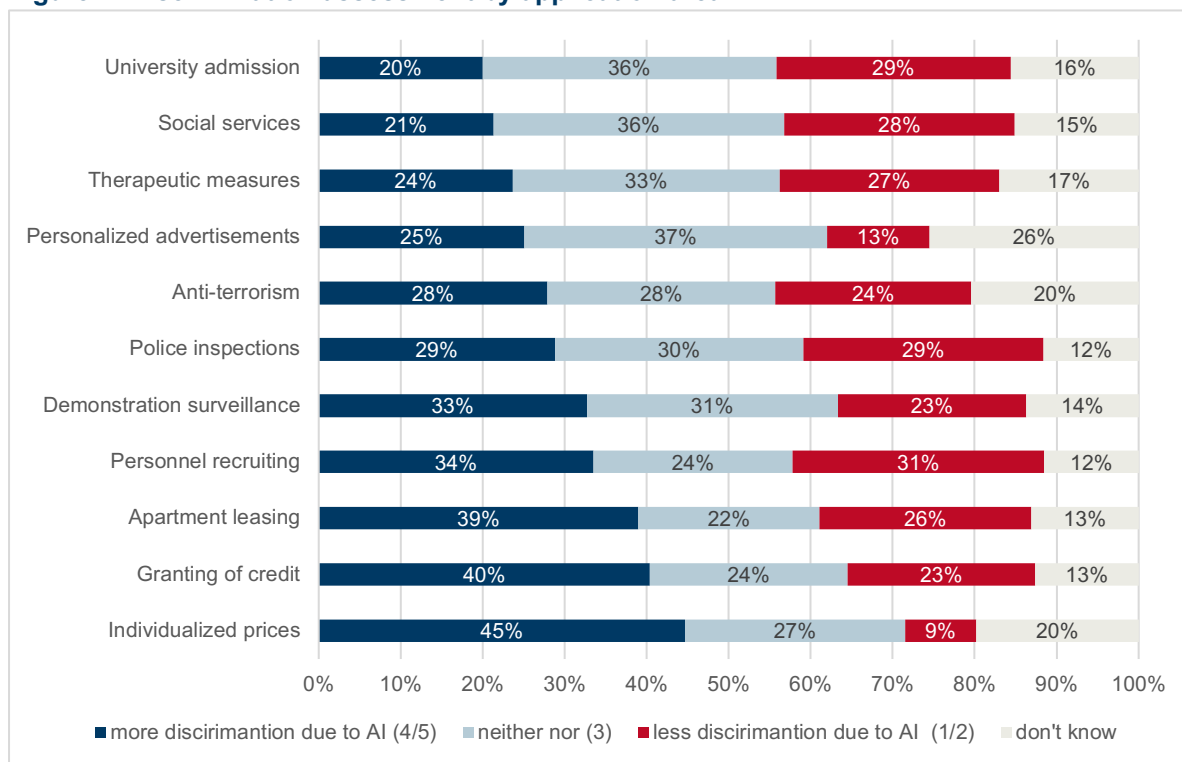
For our survey, we identified relevant applications from both papers and asked citizens whether they thought the use of AI in these areas would lead to more or less discrimination. Respondents could indicate their opinion on a five-point scale, with a score of (1) representing "significantly less discrimination" from the use of AI and (5) representing "significantly more discrimination" from AI. As before, the two upper and lower response categories were combined in each case.

The results show clear differences in the expectations of the potential for discrimination through AI in the individual application areas. In seven out of eleven applications, the respondents fear that the use of AI will lead to more discrimination. Most notably, AI activities that may entail personal economic consequences, such as individualized pricing, granting of credit, or apartment rentals, are thought to have an increased potential for discrimination through AI. In some applications, however, respondents

also see potential for reduced discrimination through the use of AI. These applications relate primarily to social applications. For example, in the case of therapeutic measures in medicine, the distribution of social welfare services and the allocation of university places, the proportion of those who expect less discrimination through the use of AI predominates. The use of AI in police checks - as has recently been the subject of more frequent media discussion - is viewed in a balanced manner by the respondents; 29% think that the use of AI will lead to more or less discrimination in this area each.

In the majority of applications, AI is associated with increasing rather than decreasing discrimination. Although the overall level of awareness of the problem of AI discrimination is in a moderate range, it can be seen that in specific applications people or institutions are considered to be less discriminating than machines.

Figure 2: Discrimination assessment by application area



Annotation: N=905

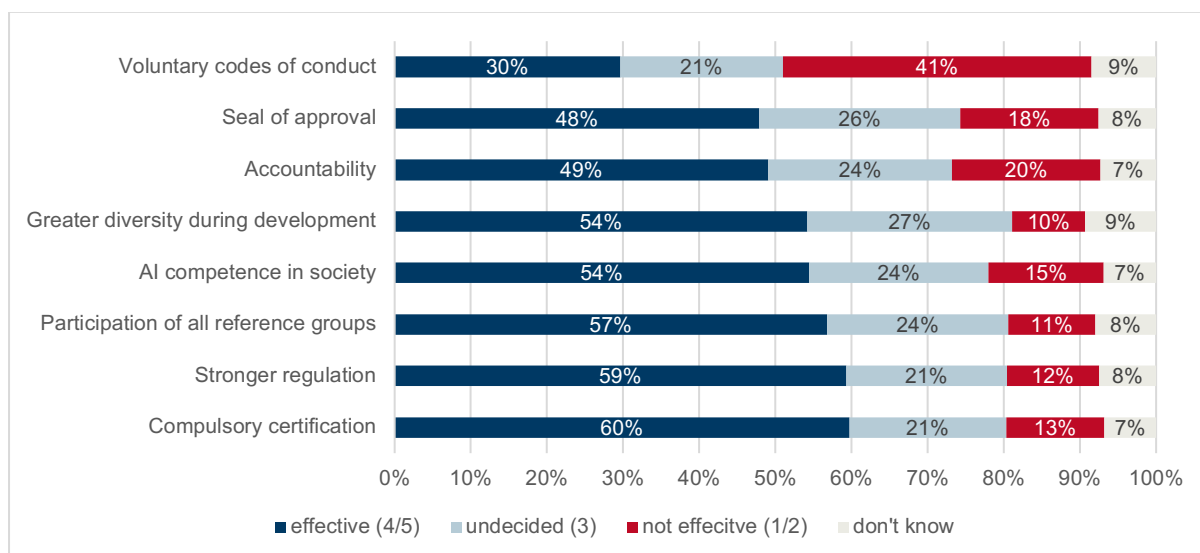
Question: What do you think: will AI lead to more or less discrimination in the following areas? The term discrimination means that a person or a group is treated preferentially or disadvantageously in comparison to others because of particular personal characteristics. (1=significantly less discrimination due to AI; 5=significantly more discrimination due to A)

## Respondents favor regulation to prevent AI discrimination

In its guidelines, the EU Commission has proposed a number of measures to prevent discrimination from AI applications. These include economic interventions (e.g., through regulation), as well as social initiatives (e.g., through education). But how effective do citizens think these are? The assessment of citizens is fundamental in that public opinion can have significant political weight in initiating, accelerating or, conversely, preventing the implementation of policies. Again, citizens were able to indicate their assessment of the effectiveness of the countermeasures on a five-point scale; for the visualization of the results, the answer options (1) and (2) were combined into "(rather) not effective" and (4) and (5) into "(very) effective".

The first observation is that many of the measures proposed by the EU to combat discrimination through AI are considered to be effective. First and foremost, regulatory interventions (e.g., compulsory certification or stronger government regulation) are seen as highly effective. However, strengthening competence in AI and greater involvement of diverse populations in development and implementation are also considered effective. The lowest effectiveness is attributed to the establishment of voluntary codes of conduct by entrepreneurs. Finally, respondents were asked to indicate which of the countermeasures mentioned should be implemented most urgently. Respondents were able to name up to three countermeasures.

Figure 3: Assessing the effectiveness of discrimination countermeasures



Annotation: N=898

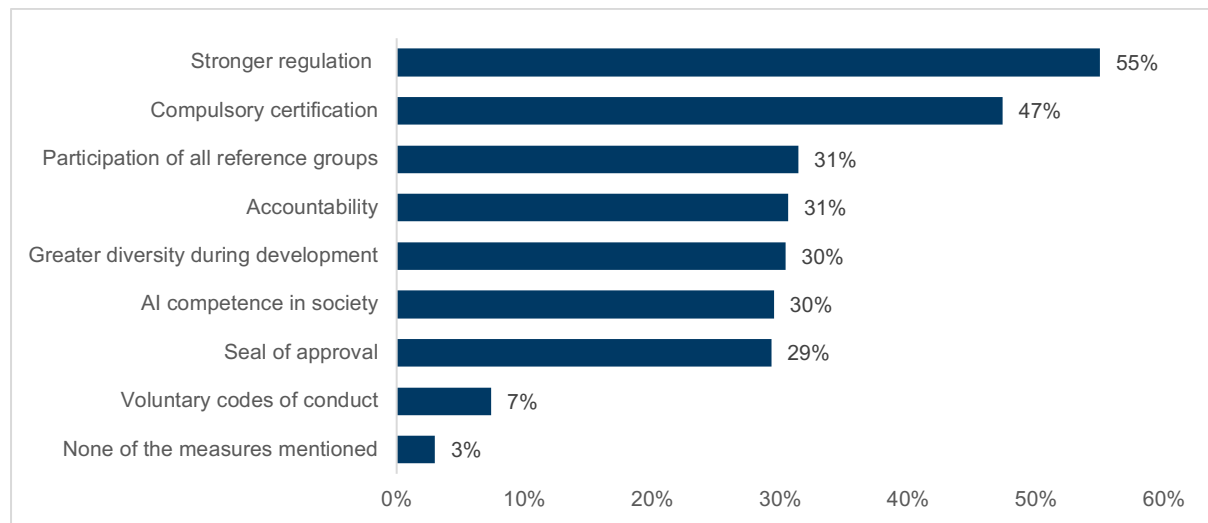
Question: Various options are being discussed for making AI systems less discriminatory. How effective do you consider the following measures to be in reducing possible discrimination - even if you do not presently see any danger from AI yourself? (1=not effective; 5= very effective)

The evaluation of the prioritization of countermeasures shows that regulatory interventions are likewise the most popular here as well. Social factors and voluntary controls are perceived as being of only limited urgency. Only three percent of respondents do not consider any of the measures mentioned to be useful.

Overall, a remarkable picture emerges in the assessment of measures to combat discrimination through AI. Respondents are in favor of stronger regulation and certification of AI technology to prevent discrimination. This can be interpreted as a sign that the European way of regulating AI with binding regulations and not

relying solely on measures from industry is seen as desirable by German citizens.

**Figure 4: Prioritization of discrimination countermeasures**



Annotation: N=898

Question: And which of the measures do you think should definitely be implemented? You can select up to three measures. If none of the measures seem necessary to you, then you can also select the answer "None of the measures mentioned".

## Conclusion

The survey results provide a contribution to the current discussion regarding discrimination through AI technologies. Overall, it is clear that among the respondents there is a rather moderate awareness of the problem of AI discrimination. However, when citizens are asked directly whether the use of AI leads to more or less discrimination in individual application areas, a majority of applications are expected to be more discriminatory - especially in areas that entail individual financial consequences. An application of AI in the social and educational sectors, however, is more likely to be associated with less discrimination. Finally, the majority of the German population, which has a basic un-

derstanding of AI, considers most of the proposed EU guidelines to be effective in preventing discrimination through AI. Interestingly, the measures that respondents feel should be implemented most urgently are those that call for greater regulation of technology and businesses. The fundamental willingness for the path to human-centered and public welfare-oriented AI, laid out by the European Union, is thus quite discernible - even though awareness of discriminatory effects of AI systems is still relatively low.

## References

Antidiskriminierungsstelle des Bundes (2017). Diskriminierung in Deutschland. Dritter Gemeinsamer Bericht der Antidiskriminierungsstelle des Bundes und der in ihrem Zuständigkeitsbereich betroffenen Beauftragten der Bundesregierung und des Deutschen Bundestages. Available at <https://www.antidis->

[kriminierungsstelle.de/SharedDocs/Downloads/DE/publikationen/BT\\_Bericht/gemeinsamer\\_bericht\\_dritter\\_2017.pdf?jsessionid=C94A23AA6C3C7B55A681617F2B372008.1\\_cid351?\\_\\_blob=publicationFile&v=16](https://www.kriminierungsstelle.de/SharedDocs/Downloads/DE/publikationen/BT_Bericht/gemeinsamer_bericht_dritter_2017.pdf?jsessionid=C94A23AA6C3C7B55A681617F2B372008.1_cid351?__blob=publicationFile&v=16)

Beck, S., Grundwald, A., Jacob, K., & Matzner, T. (2019). Künstliche Intelligenz und Diskriminierung. *Whitepaper of Plattform Lernende Systeme*. Available at [https://www.plattform-lernende-systeme.de/publikationen-details/kuenstliche-intelligenz-und-diskriminierung-herausforderungen-und-loesungsansaeetze.html?file=files/Downloads/Publikationen/AG3\\_Whitepaper\\_250619.pdf](https://www.plattform-lernende-systeme.de/publikationen-details/kuenstliche-intelligenz-und-diskriminierung-herausforderungen-und-loesungsansaeetze.html?file=files/Downloads/Publikationen/AG3_Whitepaper_250619.pdf)

Borgesius, F.Z. (2018). Discrimination, artificial intelligence, and algorithmic decision-making. *Council of Europe*. Available at <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>

European Commission (2020): White Paper on Artificial Intelligence - A European approach to excellence and trust. Available at [https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020\\_en.pdf](https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf)

To be cited as: Kieslich, K., Starke, C., Došenović, P., Keller, B., & Marcinkowski, F. (August, 2020). Artificial Intelligence and Discrimination. *Factsheet Nr. 2 of the Meinungsmonitor Künstliche Intelligenz. [Opinion Monitor Artificial Intelligence]*. Available at <https://www.cais-research.de/wp-content/uploads/Factsheet-2-AI-Discrimination.pdf>

**Contact:**

Kimon Kieslich, M. A.

**Email:**

memoki@hhu.de

**Website:**

www.cais.nrw/memoki