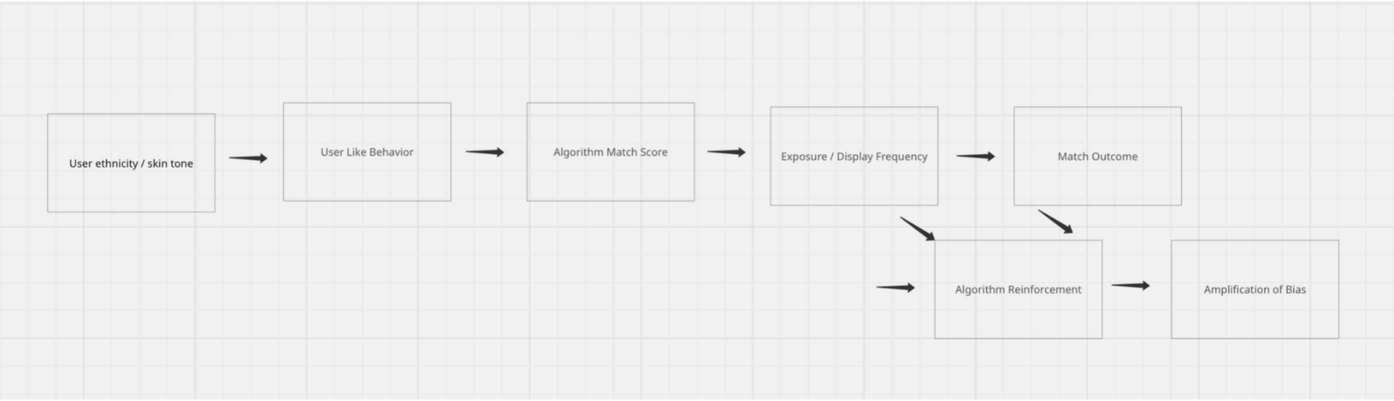
1. First Impression: Ethical Issues & Origins

The key ethical issue is algorithmic discrimination / bias: users with darker skin tones or non-Dutch backgrounds seem to be shown fewer potential matches, reducing their opportunities. Breeze's algorithm reinforces user preferences (which may already be biased) and amplifies disparities. Because the matching algorithm is self-learning and opaque, Breeze doesn't fully understand how "like behavior" feeds into match probabilities, and thus it fails to detect and correct discriminatory effects. Even if the preferences are "natural" or emergent, the system has a duty to prevent indirect discrimination.

Finally, there is a tension with data protection (AVG / GDPR): correcting bias may require using sensitive attributes (skin color, ethnicity), which are restricted under privacy law. And making the model biased makes it also vulnerable for inefficiencies like still promoting darker skin tone profiles that are not active etc (features that push a profile).

2. DAG

DAG for the dilemma:



- User Ethnicity / SkinTone influences User Like Behavior (preferences).

- User Like Behavior feeds into Algorithm Match Score (the core model).

- Match Score determines Exposure / Display Frequency (how often one is shown).

- That in turn influences Match Outcome (actual matches).

- There's a feedback loop: Exposure / Display Frequency also influences future User Like Behavior, reinforcing prior patterns (Algorithm Reinforcement / Amplification).

3. Reevaluation after Drawing the DAG

After mapping the DAG, some aspects I might have missed in my first impression:

- Bias grows over time: it doesn't stay the same but strengthens itself through feedback loops.

- User preferences aren't neutral: they are shaped by social norms and prejudice.

- The algorithm is a black box: Breeze doesn't fully know how it makes decisions, which makes fixing discrimination hard.

4. Recommendation to a Data Scientist

So how can Breeze improve its model or future algorithms?

- Use causal diagrams (DAGs) to show where bias comes in and how it grows, and find spots where you can intervene (Creager et al., 2019).

- Check bias with sensitive attributes (like ethnicity or skin tone) for auditing, even if you don't use them in the model. If possible, use privacy-friendly methods (Toreini et al., 2023).

- Keep people involved: make sure users can understand or question algorithm decisions, but do not put a human in the loop (Glickman & Sharot, 2024). People do not improve model bias by using human intervention. Rather do it via model design, bias detection methods etc.

- Monitor regularly: because bias can increase over time, keep auditing and updating the system.