

Aanbevelingen van een computer – hoe betrouwbaar zijn ze?

Recommender systems (letterlijk vertaald 'aanbevelingssystemen') zijn de drijvende kracht achter aanbevelingen en feeds in onder andere webwinkels, streamingdiensten en sociale media. Deze aanbevelingen hebben grote invloed op welke artikelen we lezen of kopen, of welke meningen we horen. Echter, deze aanbevelingen zijn vaak niet objectief, waardoor bestaande ongelijkheden in de samenleving kunnen vergroten. In dit artikel beschrijven we hoe transparante algoritmes en verklaringen dit tegen kunnen gaan, en presenteren we interactiepatronen die gebruikers meer controle over hun aanbevelingen geven.

Karlijn Dinnissen, Hanna Hauptmann, Eelco Herder, Judith Masthoff en Aletta Smits

In veel online platforms kunnen gebruikers kiezen uit een grote hoeveelheid zaken, zoals artikelen in een webshop, posts op sociale media, of series of muziek bij streamingdiensten. De platforms proberen hun gebruikers daarom te helpen met gepersonaliseerde aanbevelingen. We laten bijvoorbeeld streamingdiensten muziek voor ons afspelen zonder elk nummer zelf te kiezen. De sociale media stellen ook voor het grootste deel zelf onze tijdlijn samen: een door een algoritme samengestelde selectie van berichten, op basis van informatie die het platform over ons heeft. De drijvende kracht achter aanbevelingen op het internet zijn *recommender systems* (RS). Dit zijn algoritmes die boeken, films, nieuwsartikelen en andere (online) producten voor een gebruiker selecteren. Die selectie kan plaatsvinden door het product op basis van de inhoud te koppelen aan gebruikerswensen (*content-based recommendation*) of door producten te identificeren die gebruikers met een vergelijkbare smaak hebben gekozen (*collaborative filtering*) (Ricci et al., 2010). Met name die laatste techniek wordt vaak toegepast, omdat keuzes uit het verleden een goede indicatie zijn van producten die gebruikers in de toekomst waarschijnlijk aantrekkelijk of nuttig zullen vinden.

Gebruikers zien de aanbevelingen doorgaans in de vorm van een lijst of een *feed* waaruit ze simpelweg kunnen kiezen, zoals popsongs of series die ze nog niet kennen maar wel leuk zouden kunnen vinden. Daarmee neemt het algoritme de adviesfunctie over van een winkelvekoper, of van de redacteur die een programma samenstelt. Omdat het RS-algoritme bepaalt wat in de feed komt – en wat niet – beïnvloedt het daarmee direct onze beslissingen. Hoe meer we gebruik maken

van RS, des te belangrijker is het dat we goede interactiemogelijkheden hebben om deze systemen te kunnen gebruiken en controleren.

In dit artikel gaan we dieper in op de invloed van RS op ons dagelijks leven, en de invloed die wij zelf daarop kunnen uitoefenen. Dit doen we aan de hand van de volgende onderzoeksvragen, die elk in een eigen paragraaf worden behandeld.

- In welke mate beïnvloeden RS onze (dagelijkse) beslissingen?
- Hoe objectief en vrij van bias zijn RS? Wat betekent dit voor gebruikers en andere belanghebbenden van zo'n systeem?
- Kan transparantie en uitleg zorgen voor eerlijkere RS?
- Welke interactiepatronen kunnen gebruikers controle geven over RS?

Beslissingen nemen met behulp van RS

In toenemende mate vertrouwen we op apps, die veelal gebruik maken van RS, bij het nemen van dagelijkse beslissingen. Dit kunnen relatief onbelangrijke beslissingen zijn, zoals welk restaurant we die avond gaan bezoeken, maar ook onze politieke keuzes en onze nieuwsconsumptie worden deels bepaald door feeds van sociale media en gepersonaliseerde zoekresultaten (Herder et al., 2024).

In de beginjaren van het internet was het lastig om de juiste informatie of de juiste website te vinden. Dat was weliswaar frustrerend, maar juist het zoekproces en het raadplegen van verschillende bronnen en invalshoeken is nuttig om je mening te vormen. In de afgelopen twee decennia zijn zoekmachines en de algoritmes van sociale media zo volwassen geworden dat ze vrijwel altijd

relevante resultaten en suggesties laten zien, met als gevolg dat gebruikers niet verder zoeken.

Bij het bepalen wat wel of niet in de feed (of zoekresultaten) komt, treden ongewenste effecten op als filterbubbels die gebruikers in complottheorieën kunnen doen geloven, of bestaande maatschappelijke ongelijkheden versterken (Baeza-Yates, 2018), zoals de dominantie van mannelijke artiesten in de muziekwereld – meer hierover in de volgende paragraaf. Omdat RS mede bepalen welk nieuws, welke media, welke personen en welke berichten ons bereiken en welke niet, oefenen ze een macht uit die directe gevolgen heeft voor onze samenleving.

RS kunnen worden geschaard onder het (veel) bredere begrip ‘artificiële intelligentie’ (AI). In de media wordt doorgaans over AI (en indirect ook RS) gesproken als algoritmes die volledig zelfstandig hun gang gaan, met allerlei ongewenste gevolgen. Dit ligt echter genuanceerder: zoals we in de volgende paragraaf beschrijven, hebben de instituten die RS inzetten zelf een verantwoordelijkheid bij het vormgeven van de algoritmes. Verder – en dat is de centrale boodschap van dit artikel – kunnen en moeten de gebruikers van RS een actieve rol spelen bij het al dan niet accepteren van aanbevelingen.

Gebruikers hebben ondersteuning nodig om te bepalen of ze zoekresultaten in twijfel moeten trekken, een productsuggestie aan willen nemen, of juist om alternatieven moeten vragen. Dat is de focus van het vervolg van dit artikel, waarbij we in de komende paragraaf beginnen met *fairness* en *bias* in RS.

Eerlijkheid en objectiviteit in RS: *fairness* en *bias*

Vaak veronderstellen gebruikers dat een algoritme ‘objectief’ keuzes maakt, zonder inmenging van enige belangen of vooroordelen. In zekere zin is dat een logische gedachte, omdat computers inderdaad geen emotie of culturele achtergrond hebben, waar dit bij mensen wel invloed heeft op hun keuzes. Echter, algoritmes worden ontwikkeld door mensen, die ook vaak werken bij een bedrijf dat bepaalde doelen en belangen heeft die impact hebben op bijvoorbeeld wat wordt aanbevolen. Daarnaast worden algoritmes getraind op basis van data die komt uit de echte wereld, en die *bias* kan bevatten die wellicht de resultaten van het algoritme oneerlijk maakt. Zulke *bias* kan ontstaan op verschillende punten, waarvan we er hier enkele zullen noemen.

Ten eerste kan ongelijkheid ontstaan tijdens het verzamelen van data. Neem bijvoorbeeld een RS dat nieuwsartikelen aanbeveelt; als dat systeem uitsluitend getraind wordt op basis van wat Nederlanders die in Amsterdam wonen eerder hebben gelezen, dan is de kans groot dat aanbevelingen niet goed aansluiten bij een Engelsman die recent is geëmigreerd naar Groningen en graag op de hoogte wil raken van het nieuws in Nederland. Het systeem werkt daardoor minder goed voor de ene gebruiker dan voor de andere.

Een andere bron van *data bias* kan een ongelijke verdeling zijn die al bestaat in de ‘echte wereld’, en die op die manier gereflecteerd wordt in de verzamelde data. Denk bijvoorbeeld aan het feit dat technische functies historisch gezien meestal bekleed werden door mannen. Als je puur die data als uitgangspunt neemt voor een algoritme dat voorspelt welke persoon het beste past bij een technische vacature, dan zou het zomaar kunnen dat het systeem alleen mannelijke kandidaten zal aanbevelen¹. Het is daarom belangrijk dat er wordt stilgestaan bij dit soort historische ongelijkheden, om te voorkomen dat die – impliciet of expliciet – worden overgenomen door RS.

Een verdere vorm van *bias*, *amplificatiebias* of ook *popularity bias*, vindt zijn oorsprong in een soort sneeuwbaaleffect: als er al veel meer items worden aanbevolen van een bepaalde groep aanbieders, worden die meer geconsumeerd, waardoor ze daarna wellicht nóg vaker aanbevolen zullen worden, enzovoort. Hierdoor kunnen artiesten die al heel populair zijn, nóg populairder worden – dit ten koste van de minder bekende artiesten.

Als onderzoekers kunnen we er zelf al een behoorlijk beeld van krijgen hoe ongelijkheid in een RS geïntroduceerd, gemeten, en ook opgelost kan worden. Hieronder geven we enkele voorbeelden. Maar omdat menselijke normen en waarden zo persoonlijk en subjectief zijn, is het erg belangrijk om direct contact te zoeken met de mensen die impact ondervinden van een bepaald systeem.

Aan de Universiteit Utrecht hebben we zulk onderzoek gedaan voor RS bij muziek-streamingdiensten. Daarvoor gingen we in gesprek met artiesten om te kijken hoe groot de rol van muziek-aanbevelingen was in hun carrière, in hoeverre het voor hen duidelijk was hoe deze systemen werken, en of zij vonden dat er op het gebied van eerlijkheid voor hen als artiest nog verbeteringen te behalen waren (Dinnissen & Bauer, 2023). Op die laatste vraag waren artiesten het er unaniem over eens dat nieuwe en onbekendere artiesten een grotere kans moeten krijgen om aanbevolen te worden – en dat dus de eerdergenoemde *amplificatiebias* moet worden tegenwerkt.

Een ander probleem dat in de muziekwereld vaak ter sprake komt is de overrepresentatie van mannelijke artiesten. De artiesten die wij spraken waren het over het algemeen eens dat hier verandering in moest komen, maar sommigen vonden dat deze verandering niet zou moeten gebeuren door algoritmes meer muziek van niet-mannelijke artiesten te laten aanbevelen, maar eerder maatschappijbreed plaats zou moeten vinden.

Ook op het gebied van transparantie gaven zowel artiesten als anderen werkzaam in de muziekindustrie aan dat er nog veel ruimte voor verbetering is. Hieronder zetten we uiteen hoe transparantie door middel van uitleg in elkaar zit en welke principes daaraan ten grondslag liggen.

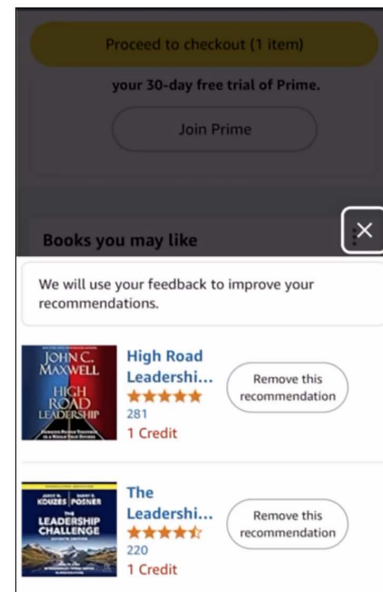
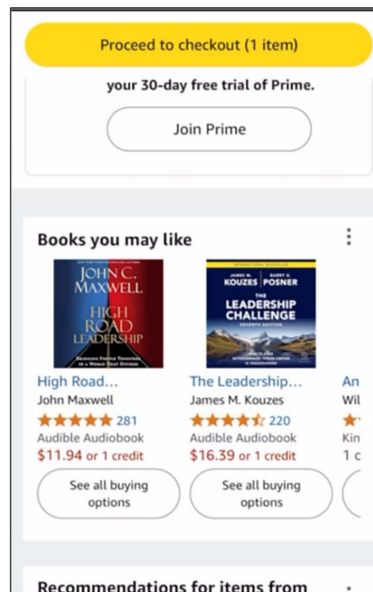
Transparantie en uitleg in RS

Om meer transparantie te bieden, wordt vaak uitleg toegevoegd aan aanbevelingen, zodat mensen begrijpen op basis van welke criteria de aanbevelingen zijn geselecteerd. Bijvoorbeeld waarom het systeem denkt dat een aanbevolen muziekstuk past bij hun smaak: “Fans van Bob Dylan zoals jij houden vaak ook van de muziek van Brandi Carlile.”

Uitleg kan echter veel meer doen dan alleen transparantie verhogen. Het kan het gemakkelijker maken om een keuze te maken, het vertrouwen in het systeem verhogen, tevredenheid bevorderen, gebruikers overhalen om een aanbeveling op te volgen, hen helpen om betere beslissingen te nemen, of controle geven. Uitleg kan gebruikers zelfs in staat stellen om in te grijpen als een RS slechte aanbevelingen selecteert op basis van verkeerde veronderstellingen (Tintarev & Masthoff, 2012). Uitleg kan op veel manieren gegeven worden – zie Tintarev & Masthoff, 2022. Afhankelijk van wat je wilt bereiken, heb je een ander soort uitleg nodig. Om gebruikers te helpen bij het selecteren van een film zijn tenslotte andere argumenten nodig dan bij het overtuigen dat een specifieke film voor hen de juiste keuze is. De wijze waarop uitleg gegeven wordt, moet vaak ook worden aangepast aan de persoon. Zo wensen sommige mensen meer detail dan anderen, hebben sommigen meer achtergrondkennis, of gaan meer af op de opinies van anderen. Uitleg kan dan – in plaats van aan te stippen wat mensen die op je lijken vinden, zoals hierboven – meer inhoudelijke argumenten geven: “Brandi Carlile’s verhalende trant lijkt op die van Bob Dylan.”

Deze vormen van uitleg geven de gebruikers inzicht in waarom een artiest als Brandi Carlile wordt aanbevolen, en kunnen daarmee de gebruiker in staat stellen aan te geven *waarom* ze de muziek van die artiest niet waarderen, door bijvoorbeeld de artiest zelf of juist de vermeende interesse in ‘verhalende songs’ expliciet weg te klikken.

Veel onderzoek naar uitlegmechanismen is erop gericht om het gebruikers naar de zin te maken of ze vertrouwen in het platform te geven. Daarnaast zijn commerciële platforms er zeker ook op gericht om gebruikers over halen tot consumptie. Echter, naast het ondersteunen van bestaande voorkeuren en gewoontes, kan uitleg ook een belangrijke rol spelen in het *veranderen* van menselijk gedrag, bijvoorbeeld door mensen bewust te maken van ongelijkheden in de muziekindustrie en ze te motiveren om vaker muziek van nieuwe, vrouwelijke artiesten een kans te geven.



Afbeelding 1. Interactie met applicatie voor meer controle op een recommender system. Links: controle over items in de lijst. Rechts: Verwijderen van aanbevolen items.

Interactiepatronen voor controle over RS

Een stap voorbij uitleg is gebruikers meer controle te geven over wat een algoritme aanbeveelt. Dat kan bijvoorbeeld door gebruikers te laten reageren op de aanbevelingen. Afbeelding 1 toont een voorbeeld van zo’n interactie. Als een gebruiker in deze webwinkel een boek koopt, krijgt hij vlak voor het betalen automatisch een lijst met ‘Books you may like’ te zien (linkerafbeelding). Daar hebben ze geen keuze in en dus geen controle over. Wel hebben ze een zekere mate van controle over welke boeken in de lijst komen te staan, omdat ze de lijst zelf kunnen aanpassen.

Deze website geeft de mogelijkheid om aanbevelingen uit die lijst te verwijderen (rechterfiguur). Het algoritme verzamelt daarmee informatie over de voorkeuren van de gebruiker. Bij een volgende aankoop wordt de ‘Books you may like’ aangepast. Voor het gevoel van de gebruiker zijn de aanbevelingen daarmee eerder een coproductie tussen algoritme en mens dan iets wat de mens door het algoritme voorgeschoteld krijgt. Deze en veel meer interacties kunnen voor gebruikers dus het gevoel van controle, transparantie en autonomie verhogen.

Over het algemeen worden er drie soorten interactie onderscheiden in de literatuur, namelijk interacties die de gebruiker helpen om de *input* waar het algoritme mee werkt aan te passen, interacties waarmee het *rekenproces* van het algoritme aangepast kan worden, en interacties die helpen om de *aanbevelingen* zelf te onderzoeken en te evalueren.

Een voorbeeld van input is een streamingdienstgebruiker die tegen het algoritme kan zeggen: ‘als je mijn *discover weekly* samenstelt, tel dan nooit de muziek mee die via mijn account tussen 6 en 8 uur wordt afgeluisterd’ (anders hebben ouders van jonge kinderen alleen Sesamstraatmuziek in hun *discover weekly* zitten). Een

voorbeeld van het aanpassen van het rekenproces is als een gebruiker tegen het algoritme van een openbaarvervoer-RS kan zeggen: 'Ik vind langere reistijd minder belangrijk dan het aantal keren overstappen.' Tenslotte is een voorbeeld van interacties die de aanbevelingen helpen evalueren de mogelijkheid om resultaten anders te sorteren, bijvoorbeeld op basis van populariteit.

Designers en data-wetenschappers ontwikkelen deze interacties gezamenlijk. Designers onderzoeken welke mate van controle de gebruiker in een specifiek RS wenst; data-wetenschappers ontwikkelen op basis daarvan een RS-algoritme dat de input van de gebruiker ontvangt en verwerkt.

Designers geven echter aan dat ze vinden dat ze te weinig van interactieve AI weten om de interacties goed te kunnen ontwerpen (Smits & van Turnhout, 2023), omdat ze vinden dat ze te weinig weten van algoritmen. Om designers te ondersteunen, ontwikkelen de Universiteit Utrecht, de Hanze Hogeschool en de Hogeschool Utrecht een bibliotheek met voorbeelden van interactiepatronen.²

Elke categorie interacties wordt geïllustreerd met rijke voorbeelden, die laten zien hoe ze voor de gebruiker hun gevoel voor controle, transparantie of autonomie veranderen. Er worden daarnaast ook voorbeelden gegeven van interactiemechanismen die een goed idee lijken, maar vooral een gevoel van *cognitive overload* veroorzaken: te veel knopjes waaraan gedraaid kan worden levert juist minder gevoel van controle op.

Een ander voorbeeld van minder controle is als mensen wel iets kunnen doen, maar de gevolgen ervan niet kunnen voorspellen. Neem bijvoorbeeld de *like*-button in een bekende streamingdienst: gebruikers kunnen zeggen dat ze een serie goed vonden, maar ze kunnen niet zeggen wat ze er precies goed aan vonden (afbeelding 2). Stel dat op de streamingdienst de serie *Heartstopper* een *like* krijgt omdat een acteur zo leuk is, dan is het niet meteen de bedoeling dat dit een hele reeks aanbevelingen oplevert die allemaal over het thema van die show – kalverliefde – gaan.

Discussie en conclusie

In dit artikel hebben we laten zien hoe RS een essentiële rol in ons dagelijks leven spelen. RS-algoritmes zijn de drijvende kracht achter online platforms die wij gebruiken om muziek te streamen, producten te kopen, online contacten te onderhouden, of nieuws te lezen. Hoewel wij ons graag laten leiden door wat 'in de feed' aanbevolen wordt, is het aanbevelingsproces helemaal niet zo waardevrij als veel mensen denken. Zo worden populaire keuzes en items doorgaans extra vaak aanbevolen, en ook andere bestaande ongelijkheden worden versterkt. Om deze effecten tegen te gaan, zijn transparantie en uitleg een goed begin, omdat gebruikers hiermee meer inzicht kunnen krijgen in hoe algoritmes werken en op basis waarvan aanbevelingen worden gedaan. Dit is echter niet afdoende: het is



Afbeelding 2. Voorbeeld van een *like*-button in een applicatie waarmee invloed wordt uitgeoefend op items die worden aanbevolen.

belangrijk dat de gebruiker zelf controle kan uitoefenen, zodat die zelf in kan grijpen als een RS ondermaatse aanbevelingen geeft. Daarom wordt er gebouwd aan een bibliotheek met interactiepatronen die dit mogelijk maken.

Hoe beter de interacties ontworpen zijn, des te tevredener zijn gebruikers met de aanbevolen resultaten en des te groter is de kans dat ze een platform vertrouwen – of dat nu een muziek-streamingdienst is of een persoonlijke assistent die aanbevelingen geeft over een gezondere levensstijl. En dus ook dat ze dat platform blijven gebruiken. Dat is essentieel in een wereld waarin we steeds meer adviezen uitbesteden aan AI. Want in die wereld moet de mens ook controle houden, en blijven voelen dat ze wel degelijk invloed hebben op de adviezen waar ze aan blootgesteld worden.

Abstract

Web platforms, including online stores, streaming platforms and social media, make use of recommender systems in order to create item suggestions and item feeds. These recommendations have a strong impact on which articles we read, which items we buy, or which opinions we hear. However, recommender systems are not completely neutral and may even increase existing bias within society. This article discusses how transparency and explanations can combat such unwanted effects and presents interaction mechanisms that allow users to actively manage their recommendations.

Referenties

- Baeza-Yates, R. (2018). Bias on the web. *Communications of the ACM*, 61(6), 54–61. <https://doi.org/10.1145/3209581>
- Dinnissen, K., & Bauer, C. (2023). Amplifying artists' voices: Item provider perspectives on influence and fairness of music streaming platforms. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization (UMAP '23)* (pp. 238–249). ACM. <https://doi.org/10.1145/3565472.3592960>
- Herder, E., Stojko, L., Strecker, J., Neumayr, T., Yigitbas, E., & Augstein, M. (2024). Towards new realities: Implications of personalized online layers in our daily lives. *i-com*, 23(2), 221–229. <https://doi.org/10.1515/icom-2024-0017>
- Ricci, F., Rokach, L., & Shapira, B. (2010). Introduction to recommender systems handbook. In *Recommender Systems Handbook* (pp. 1–35). Springer. https://doi.org/10.1007/978-0-387-85820-3_1
- Smits, A., & van Turnhout, K. (2023). Towards a practice-led research agenda for user interface design of recommender systems. In L. A. (Ed.), *Human-Computer Interaction – INTERACT 2023* (Vol. 14144, pp. 127–143). Springer. https://doi.org/10.1007/978-3-031-42286-7_10
- Tintarev, N., & Masthoff, J. (2022). Beyond explaining single item recommendations. In *Recommender Systems Handbook* (pp. 763–796). Springer. https://doi.org/10.1007/978-1-0716-2197-4_19

Bijdrage aan het HF-kennisdomein

Dit artikel draagt bij aan het HF-kennisdomein door een overzicht te geven hoe recommender systems (RS) impact hebben op het gedrag en de keuzes van gebruikers. Omdat RS zijn geïntegreerd in vele online platforms die ons dagelijkse leven bepalen – zoals sociale media en streamingdiensten – is een *systeemaanpak* van belang, waarbij niet alleen de gevolgen voor individuele gebruikers worden afgewogen, maar vooral ook de effecten op de hele samenleving.

We geven een toegankelijke, systematische inleiding in de functionaliteit van RS en de oorzaken van ongewenste effecten. Als gebruikers vervolgens blind vertrouwen op de aanbevelingen die ze krijgen, vergroot dat de macht van de platforms die de keuzes aanbieden. Vervolgens wordt een *ontwerpgedreven benadering* gekozen: we geven concrete voorbeelden welke interactiepatronen transparantie bieden en de gebruikers controle geven over de werking van het aanbevelingsproces.

De rijke voorbeelden van interactiepatronen laten duidelijk zien welke invloed de *systeemprestaties* van RS hebben op het *menselijk welzijn*. Transparantie en controlemechanismen hebben een positieve invloed op de gebruikerservaring, maar vooral ook op de daadwerkelijke aanbevelingen die worden getoond en de keuzes die op basis daarvan worden gemaakt.



Eindnoten

- 1 <https://www.reuters.com/article/world/insight-amazon-scrap-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/>
- 2 Dit samen met Avans Hogeschool, de Haagsche Hogeschool, de Hogeschool Rotterdam, drie internationale universiteiten, twee designopleidingen en 15 designers.

Over de auteurs



MSc. K. (Karlijn) Dinnissen
PhD kandidaat, Universiteit Utrecht
k.dinnissen@uu.nl



Dr. H.J. (Hanna) Hauptmann
Universitair docent, Universiteit Utrecht



Dr. Ir. E. (Eelco) Herder
Universitair hoofddocent, Universiteit Utrecht



Prof. dr. ir. J.F.M. (Judith) Masthoff
Professor Interactie Technologie,
Universiteit Utrecht



Dr. A.M. (Aletta) Smits
Lector Human-Centered Technology,
Hanze Hogeschool