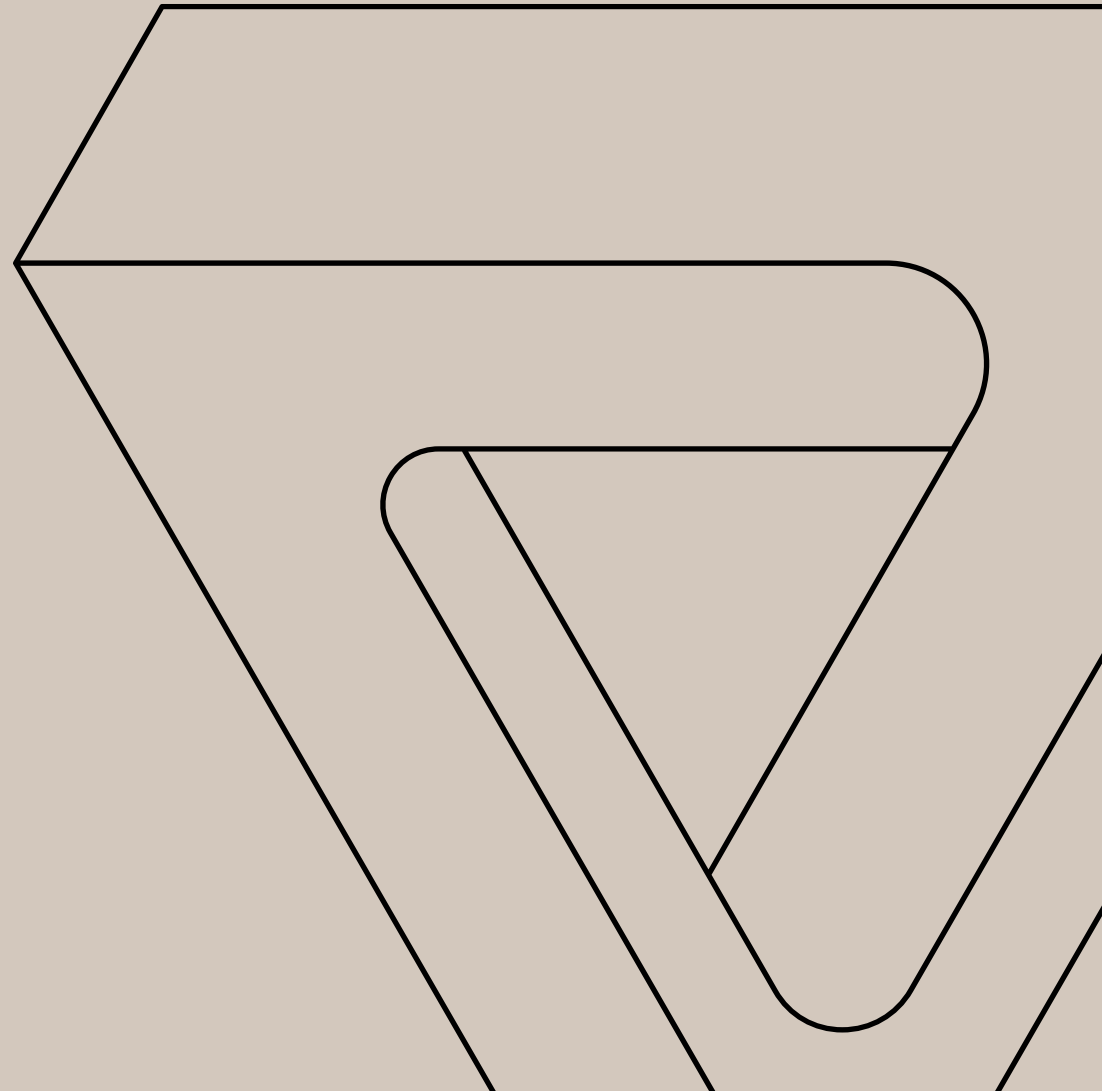


Onderzoek – Stem misleiding

Rapportage
In opdracht van het ministerie van Justitie en Veiligheid

Joyce van Brecht & Jilske Schotanus

27 mei 2024





01

CONTEXT

Voice clonen, ofwel stemklonen, is een ontwikkeling binnen de kunstmatige intelligentie die in het buitenland al regelmatig door criminelen wordt ingezet voor misleiding en oplichting. Met de campagne 'Laat je niet interneppen' wil de Rijksoverheid mensen bewust maken van online misleiding door criminelen die zich voordoen als een bekende of betrouwbare organisatie (social engineering). De campagne heeft als doel mensen te helpen deze misleiding beter te herkennen en daardoor te voorkomen dat zij slachtoffer worden van online oplichting. De Rijksoverheid verwacht dat online misleiding door stemklonen op korte termijn ook in Nederland ingezet zal worden door criminelen. Daarom wil de Rijksoverheid onderzoeken in welke mate Nederlanders stemmen die gegenereerd zijn door AI makkelijk of moeilijk kunnen herkennen en hoe ze denken over stemklonen en de risico's ervan.

02

ONDERZOEKSDOEL

Het doel van dit onderzoek is het toetsen van de hypothese dat Nederlanders een (bekende) neppe/gekloonde stem (via AI technologie) moeilijk kunnen onderscheiden van een echte stem. Daarnaast worden inzichten opgehaald over de mate waarin Nederlanders kennis en ervaring hebben met gekloonde stemmen en in hoeverre ze zich zorgen maken over misbruik van stemklonen.

03

Onderzoeksvragen

De vragen die in dit onderzoek centraal staan zijn:

- *In hoeverre kunnen Nederlanders herkennen of een stem, afgespeeld via een digitaal apparaat, echt of gekloond/nep is?*
- *Hoeveel ervaring hebben Nederlanders met stemklonen?*
- *Wat is het kennisniveau en risicoperceptie rondom stemklonen bij Nederlanders?*

Leeswijzer

01

Respondenten hebben in de vragenlijst verschillende stemfragmenten gehoord. Deze fragmenten worden in dit rapport als volgt aangeduid:

- *Gekloond (stem)fragment*: geluidsfragment dat is gemaakt met behulp van AI stemkloon-technologie.
- *Echt (stem)fragment BN-er*: geluidsfragment dat is ingesproken door radiopresentator Ruud de Wild.

02

De data is op totaal niveau gewogen naar de representatieve cijfers van Nederlanders op geslacht, leeftijd (18+), opleidingsniveau en regio. Dit betekent dat sommige antwoorden zwaarder meetellen omdat ze een hogere weegfactor hebben en anderen juist minder zwaar omdat ze een lagere weegfactor hebben. De basis (n=...) die gerapporteerd wordt, is altijd de ongewogen basis, dus het daadwerkelijke aantal mensen in die groep. De percentages slaan dus niet altijd op de basis die staat vermeld.

03

Uit een checkvraag blijkt dat 9 op de 10 respondenten de stem van Ruud de Wild (weleens tot vaak) gehoord heeft. Uit de analyses blijkt dat er geen verschillen zijn tussen de antwoorden van respondenten die zijn stem ooit gehoord hebben en respondenten die dat niet hebben. Alle respondenten zijn dan ook meegenomen in de analyses en rapportage.

04

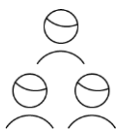
Verschillen tussen leeftijd, geslacht en opleidingsniveau zijn bekeken. Indien de verschillen tussen groepen relevant en groter dan 10% zijn, worden deze in het rapport beschreven.

Onderzoeksverantwoording



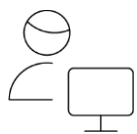
Doelgroep

Nederlanders boven de 18 jaar.



Steekproefbron

De steekproef is getrokken uit het respondentenpanel NIPObase.



Respons

In totaal hebben n=1028 respondenten de vragenlijst ingevuld. De steekproef is vergeleken met de daadwerkelijke samenstelling van de doelgroep op de kenmerken geslacht, leeftijd, opleiding en regio. Voor afwijkingen is gecorrigeerd door middel van herweging. De uitkomsten kunnen hiermee als representatief worden beschouwd voor de groep Nederlanders boven de 18 jaar.



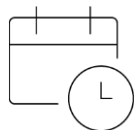
Online methode

Voor dit onderzoek is gebruik gemaakt van een online vragenlijst (CAWI: computer assisted web interviewing). Respondenten ontvingen een e-mail met daarin algemene informatie over het onderzoek en een link naar de online vragenlijst. Met een checkvraag na de eerste geluidsfragmenten is gecontroleerd of de respondenten de fragmenten goed konden beluisteren. Was dit niet het geval, dan werden ze uit de vragenlijst geleid.



Vragenlijst

De vragenlijst is opgezet door de opdrachtgever en in samenspraak met Verian definitief gemaakt. De vragenlijst bestond uit drie blokken, waarin respondenten geluidsfragmenten te horen kregen en vragen moesten beantwoorden. De invulduur bedroeg ca. 5 minuten.



Veldwerkperiode

Het veldwerk heeft gelopen van maandag 6 t/m zondag 12 mei 2024.

Samenvatting

Nederlanders hebben moeite met het herkennen van gekloonde stemfragmenten

In het onderzoek zijn verschillende stemfragmenten aan Nederlanders voorgelegd, met de vraag of het geluidsfragment de echte stem van een bekende radiopresentator of een door AI gegenereerde stem was. Het onderzoek laat zien dat Nederlanders veel moeite hebben met het herkennen van de 'echtheid' van stemmen.

Het onderzoek bestond uit drie onderdelen:

1. Herkenning gekloonde stemfragmenten
2. Identificeren echtheid stemfragmenten
3. Vragen over ervaringen, kennis en risicoperceptie rondom stemmisleiding

In het eerste blok kreeg men drie geluidsfragmenten te horen die door AI waren gegenereerd, gevolgd door de vraag of dit een gekloonde stem of echte stem was. De meeste mensen hadden het niet goed: 6 op de 10 mensen dachten dat de gekloonde stemfragmenten echt door de BN-er waren ingesproken. 1 op de 10 gaf zelfs aan dit zeker te weten.

In het tweede blok kreeg men twee tekstueel identieke fragmenten, waarvan één ingesproken door de bekende radiopresentator en één gegenereerd door AI. Ook hier hadden Nederlanders veel moeite met het herkennen van de echte en gekloonde stem: 51% identificeerde de gekloonde stem ten onrechte als echt en 49% identificeerde de echte stem ten onrechte als de gekloonde stem.

Nederlanders zijn zich hier zelf ook van bewust: slechts 2% geeft aan het *niet* lastig te vinden om te onderscheiden of een stem echt of gekloond is.

Uit het onderzoek komen verder geen duidelijke indicatoren naar voren voor het

juist herkennen van echte of gekloonde stemmen. Wel lijken de vloeiendheid en emotionele toon van de stem een rol te spelen.

Veel Nederlanders nog niet goed bekend met stemklonen

Circa een kwart van de Nederlanders (26%) is niet bekend met stemklonen. Ongeveer een derde (30%) heeft er weleens van gehoord maar weet niet precies wat het inhoudt. De rest (44%) heeft er wel over gehoord en weet wat het inhoudt.

Ondanks dat veel Nederlanders de risico's van stemklonen zien, worden er nog weinig maatregelen genomen

Circa de helft van de Nederlanders (54%) ziet het grootste gevaar van stemklonen in het namaken van stemmen waardoor het lijkt alsof je iets gezegd hebt terwijl dit in werkelijkheid niet zo is. Maar ook voordoen als een bekende met stemklonen (45%) en de schending van privacy of identiteitsdiefstal (43%) worden veel genoemd.

Slechts 2% van de Nederlanders denkt dat stemklonen in de toekomst *niet* zal worden gebruikt om mensen te bedriegen. Meer dan de helft (58%) maakt zich zorgen dat ze in de toekomst zullen worden opgelicht door middel van stemklonen.

Ondanks de zorgen is er nog weinig kennis over mogelijk te nemen maatregelen: slechts 7% weet (enigszins) welke maatregelen tegen stemklonen beschermen, 75% weet het niet.

An aerial night photograph of a long, curved bridge with warm yellow lights along its length, spanning across a body of water. The water reflects the lights and the surrounding environment, creating a shimmering effect. The bridge's structure is visible, with supports and railings. The overall color palette is dominated by deep blues and greens, contrasted with the bright yellow of the bridge lights.

02.1 RESULTATEN

Herkenning gekloonde stemfragmenten

Opzet blok 1. Herkenning gekloonde stemfragmenten

In het eerste vragenblok kregen alle respondenten drie verschillende stemfragmenten met de door AI gekloonde stem van de bekende Nederlander (radiopresentator Ruud de Wild) te horen. De respondenten kregen hier vervolgens vragen over.

Alle respondenten
n=1028

3x



Gekloonde stem - geluidsfragment 1*



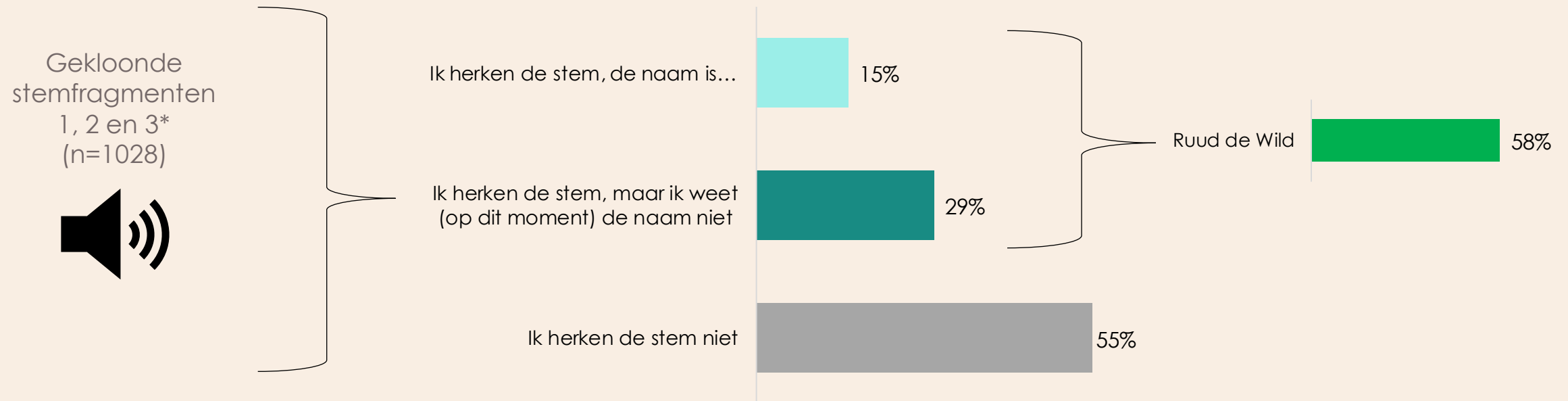
Gekloonde stem - geluidsfragment 2*



Gekloonde stem - geluidsfragment 3*

Herkennen stem in gekloonde stemfragmenten

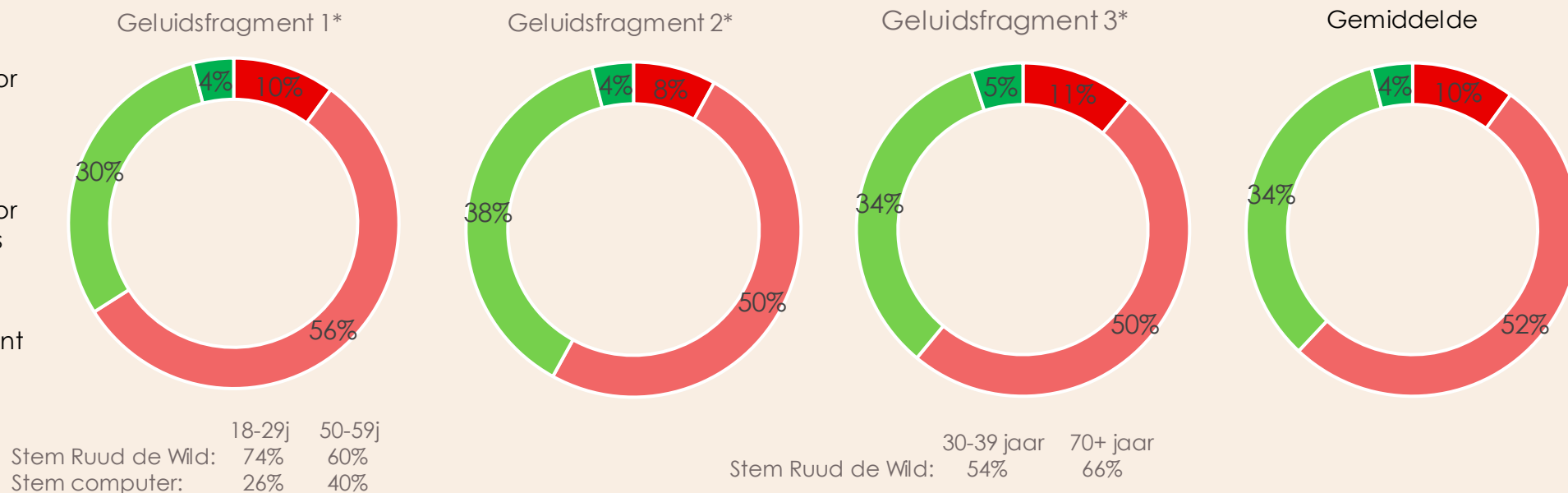
58% van de mensen die aangeeft de stem te herkennen denkt dat het de stem van de bekende Nederlander (radiopresentator Ruud de Wild) is.



Herkennen stem in gekloonde stemfragmenten

Veel mensen hebben moeite met het herkennen van de gekloonde stemfragmenten als nep: circa 6 op de 10 mensen geeft aan te denken dat de gekloonde stem echt is en door de BN-er is ingesproken. 1 op de 10 mensen geeft aan dit zeker te weten.

- Ik weet zeker dat dit fragment door Ruud de Wild zelf is ingesproken
- Ik denk dat dit fragment door Ruud de Wild zelf is ingesproken
- Ik denk dat dit fragment door een computerprogramma is gemaakt
- Ik weet zeker dat dit fragment door een computerprogramma is gemaakt



An aerial night photograph of a long, curved bridge spanning across a body of water. The bridge is illuminated with warm yellow lights, creating a glowing path against the dark blue water. The surrounding landscape is dark and textured, possibly showing some vegetation or rocky terrain. The overall mood is serene and mysterious.

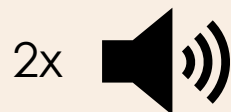
02.2 RESULTATEN

Identificeren echtheid stemfragmenten

Opzet blok 2. Identificeren fragmenten echt en gekloond

In het tweede vragenblok kregen alle respondenten beknopt uitleg over stemklonen (wat is het?). Vervolgens kregen de respondenten twee stemfragmenten te horen. De helft van de respondenten kreeg twee in tekst identieke stemfragmenten te horen met een positieve boodschap*. Een van deze stemfragmenten was ingesproken door de bekende Nederlander (Ruud de Wild) en het andere stemfragment was de door AI gekloonde stem van de BN-er. De andere helft van de respondenten kreeg twee tekstueel identieke stemfragmenten met een negatieve boodschap*, waarvan één echt was en de andere door AI gekloond. De respondenten kregen hier vervolgens vragen over.

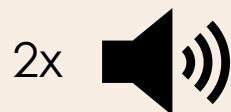
Groep 1 – positief fragment*
n = 532



Volgorde random

- Ingesproken door bekende Nederlander (echte fragment)
- Gekloond door AI (gekloonde fragment)

Groep 2 – negatief fragment*
n = 496



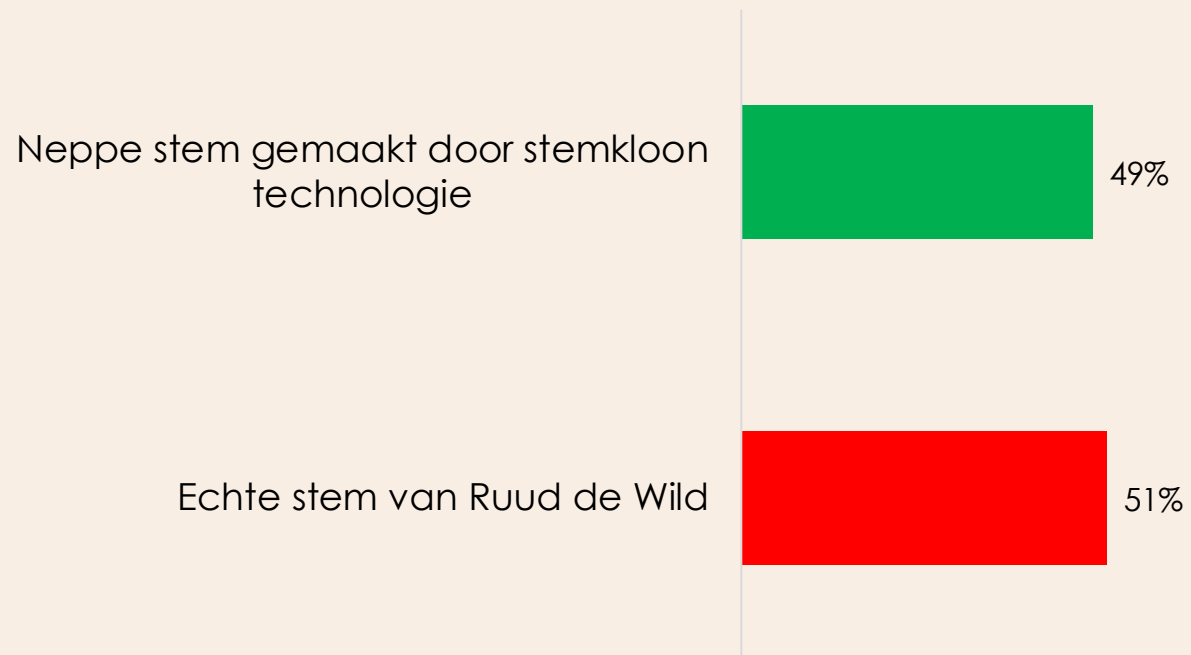
Volgorde random

- Ingesproken door Bekende Nederlander (echte fragment)
- Gekloond door AI (gekloonde fragment)

Herkenning van gekloonde stemfragmenten

Men heeft moeite met identificeren van een gekloond stemfragment: bijna de helft van de mensen denkt dat het gekloonde stemfragment echt is ingesproken door de BN-er.

Denk je dat het fragment echt of nep is?

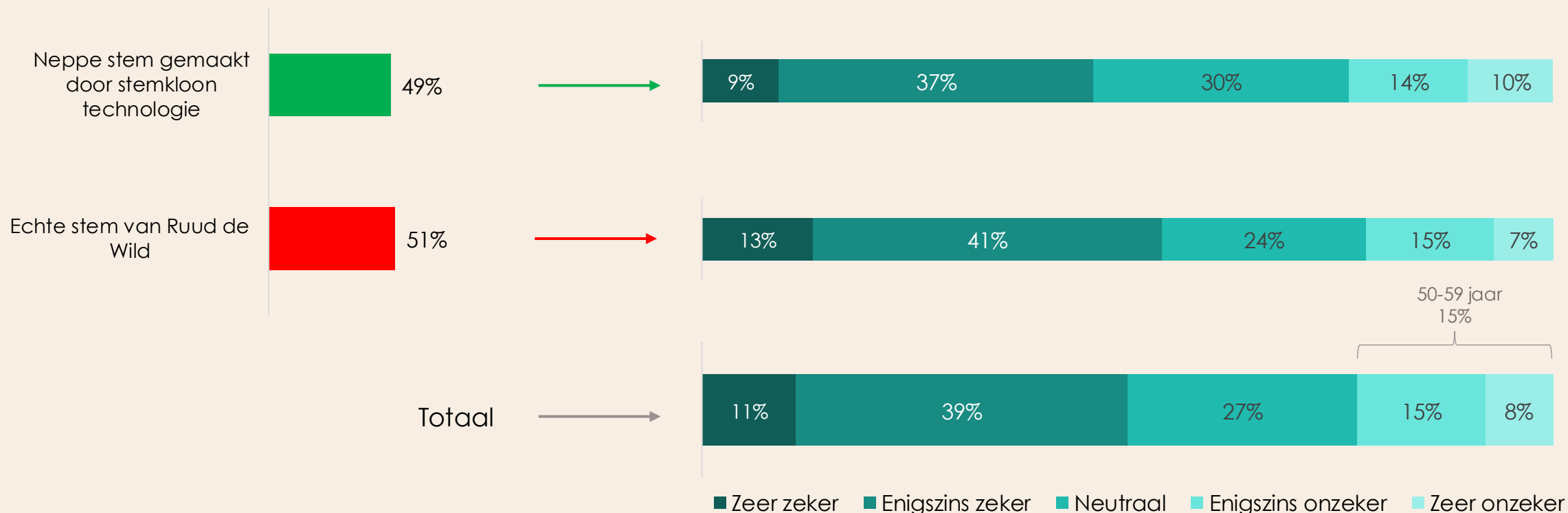


Herkenning van gekloonde stemfragmenten – zekerheid

Meer dan de helft van de mensen (54%) die het gekloonde stemfragment juist hebben herkend, geeft aan hier niet zeker over te zijn. Circa 1 op de 10 is – ten onrechte – zeer zeker dat het gekloonde stemfragment de echte stem van de BN-er is.

Denk je dat het fragment echt of nep is?

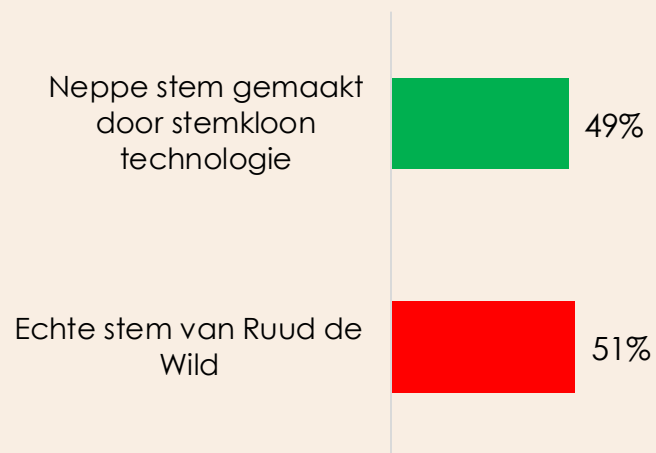
Hoe zeker ben je hiervan?



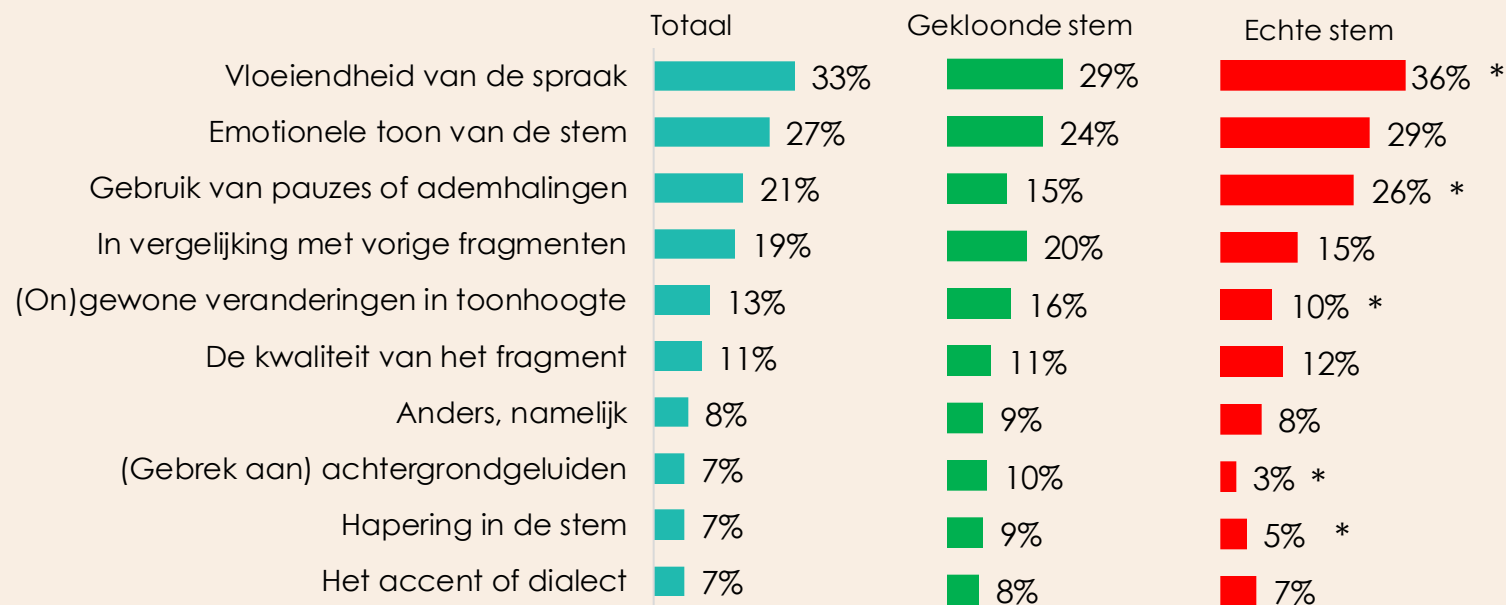
Herkenning van gekloonde stemfragmenten – redenen

Circa een derde van de mensen die de gekloonde stem als zodoende herkent geeft aan dat dit door de vloeiendheid van de spraak komt. Vloeiendheid van de spraak wordt echter ook genoemd door mensen die de gekloonde stem – ten onrechte – als echt herkennen. Er is dus geen goede voorspeller gevonden op basis waarvan respondenten konden inschatten of een fragment echt of nep was.

Denk je dat het fragment echt of nep is?



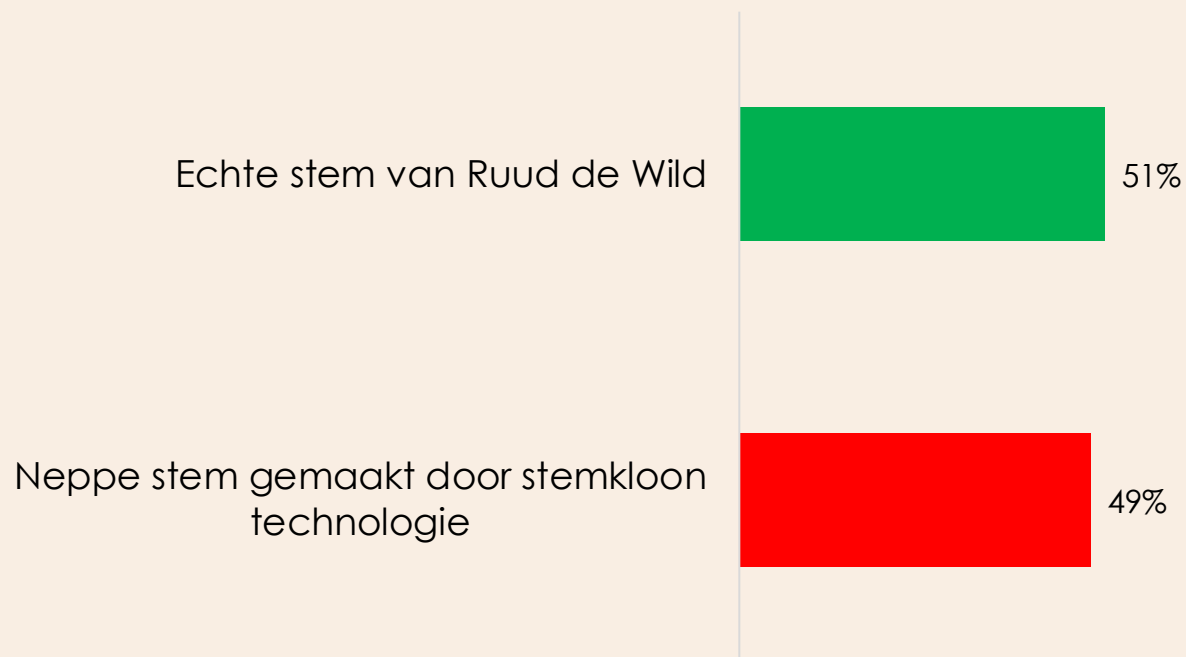
Waar leid je dit uit af?



Herkenning van echte fragmenten (stem BN-er)

Men heeft moeite met identificeren van de echtheid van een fragment: circa de helft van de mensen denkt dat het echte fragment niet ingesproken is door de BN-er maar gemaakt is met stemkloon technologie.

Denk je dat het fragment echt of nep is?



Herkenning van echte fragmenten (stem BN-er) – zekerheid

Bijna 1 op de 10 mensen geeft – ten onrechte aan – het zeer zeker te weten dat het echte stemfragment is gemaakt door stemkloon technologie.

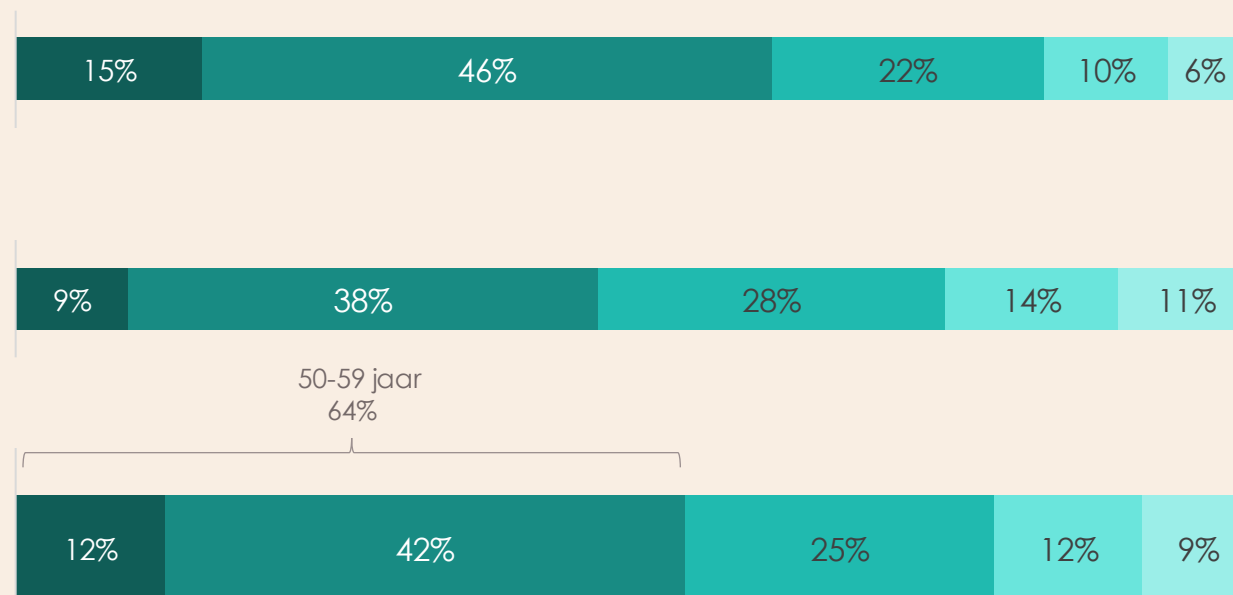
Denk je dat het fragment echt of nep is?

Echte stem van Ruud de Wild 51%

Neppe stem gemaakt door stemkloon technologie 49%

Totaal

Hoe zeker ben je hiervan?

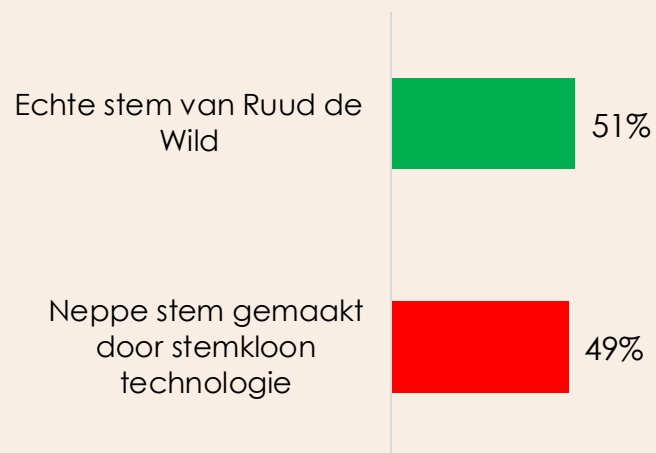


■ Zeer zeker ■ Enigszins zeker ■ Neutraal ■ Enigszins onzeker ■ Zeer onzeker

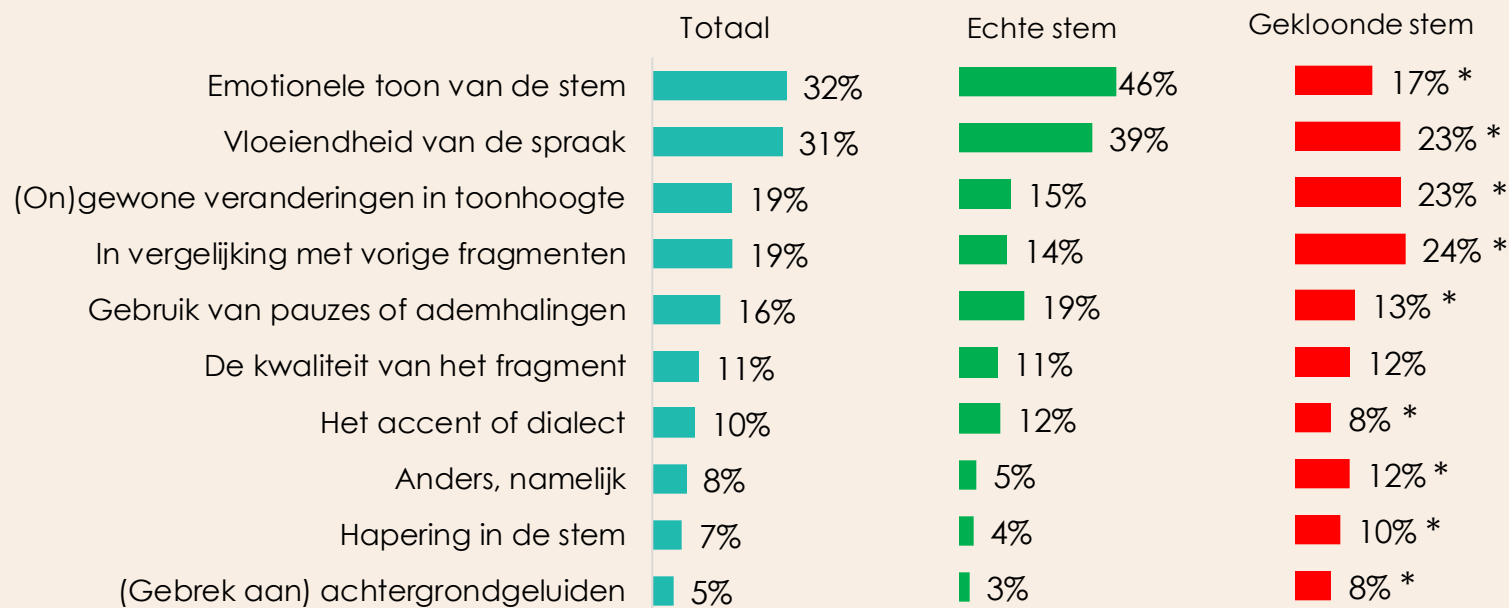
Herkenning van echte fragmenten (stem BN-er) – redenen

Bijna de helft van de mensen die de echte stem als zodoende herkent geeft aan dat dit door de emotionele toon van de stem in de fragmenten komt en 39% geeft aan door de vloeiendheid van de spraak.

Denk je dat het fragment echt of nep is?



Waar leid je dit uit af?





02.3 RESULTATEN

Ervaringen, kennis en risicoperceptie

Bekendheid met stemklonen

Circa een kwart van de Nederlanders is niet bekend met stemklonen.

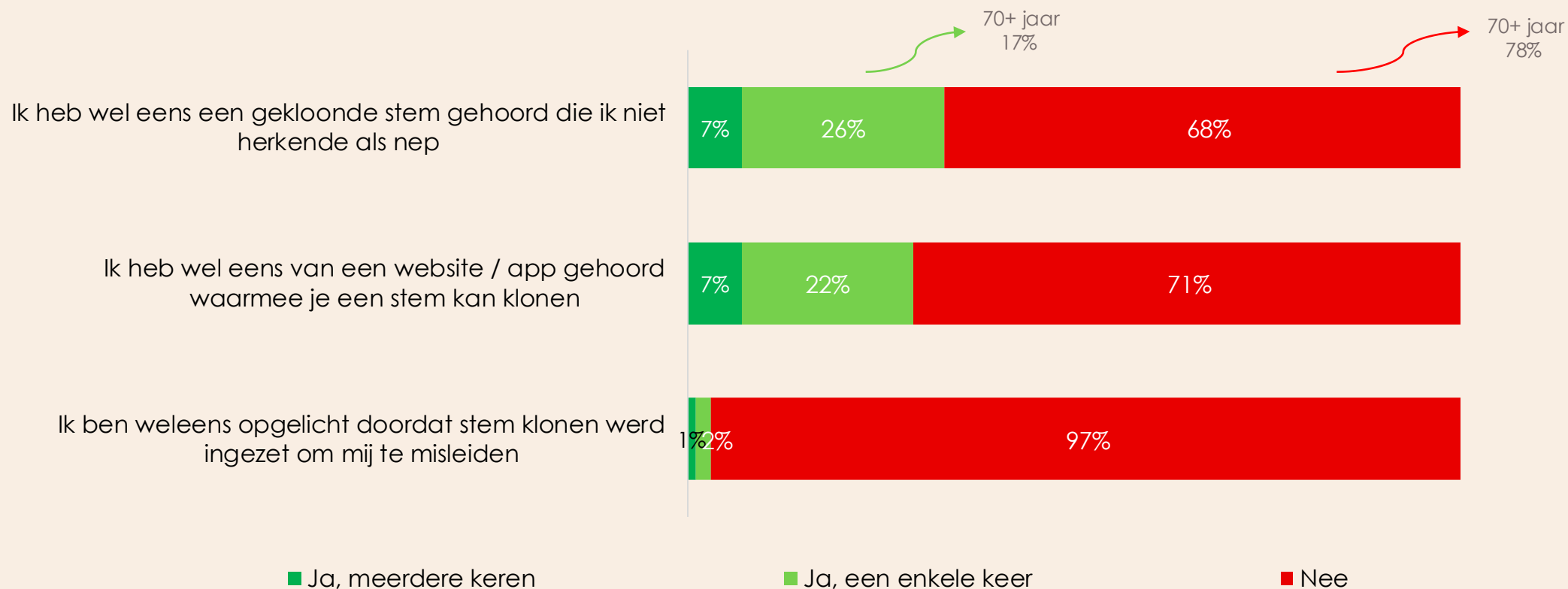
Ben je bekend met stemklonen?



- Ja, ik heb er eerder over gehoord en weet wat het inhoudt
- Ja, ik heb er wel eens van gehoord, maar weet niet precies wat het inhoudt
- Nee, ik ben er niet mee bekend

Ervaring met stemklonen

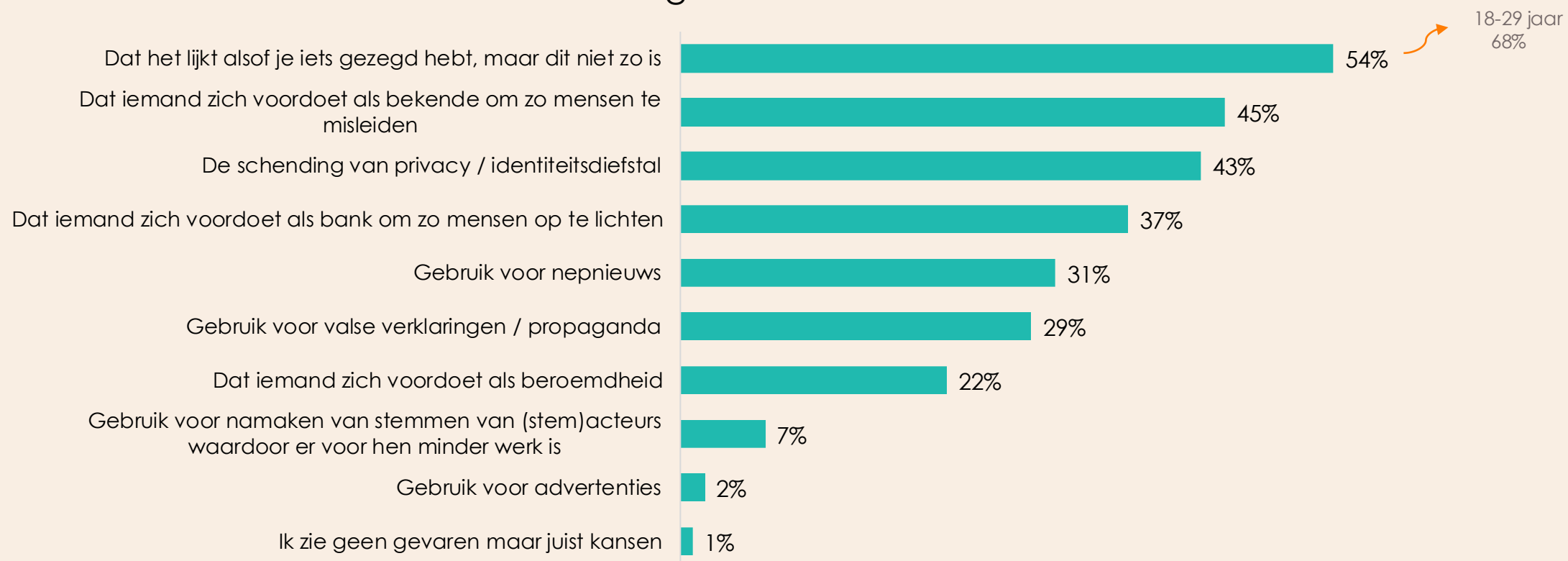
Circa een derde van de Nederlanders heeft wel eens een gekloonde stem gehoord die toen niet herkend werd als nep. 3% van de Nederlanders zegt weleens opgelicht te zijn door middel van stemklonen.



Risicoperceptie stemklonen

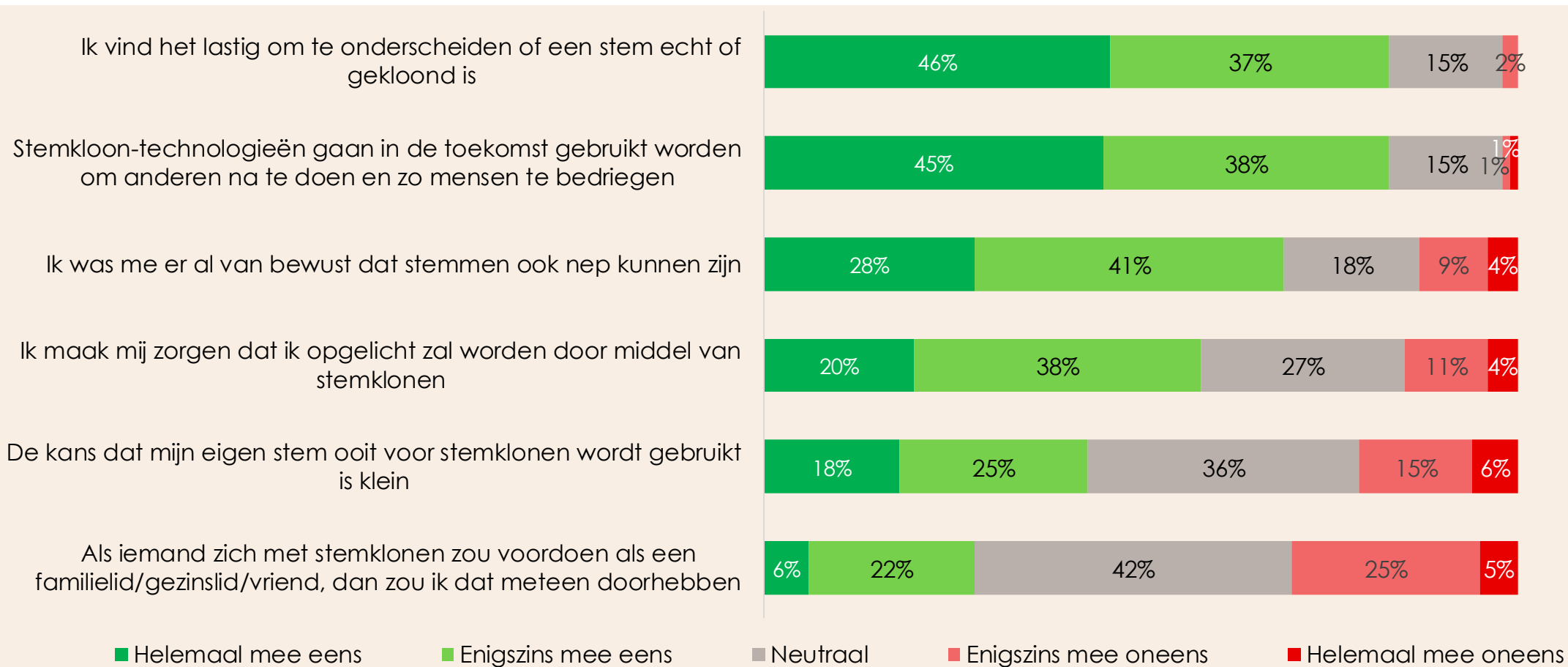
Circa de helft van de Nederlanders ziet het grootste gevaar van stemklonen in het namaken van stemmen waardoor het lijkt alsof je iets gezegd hebt terwijl dit in werkelijkheid niet zo is.

Grootste gevaar van stemklonen



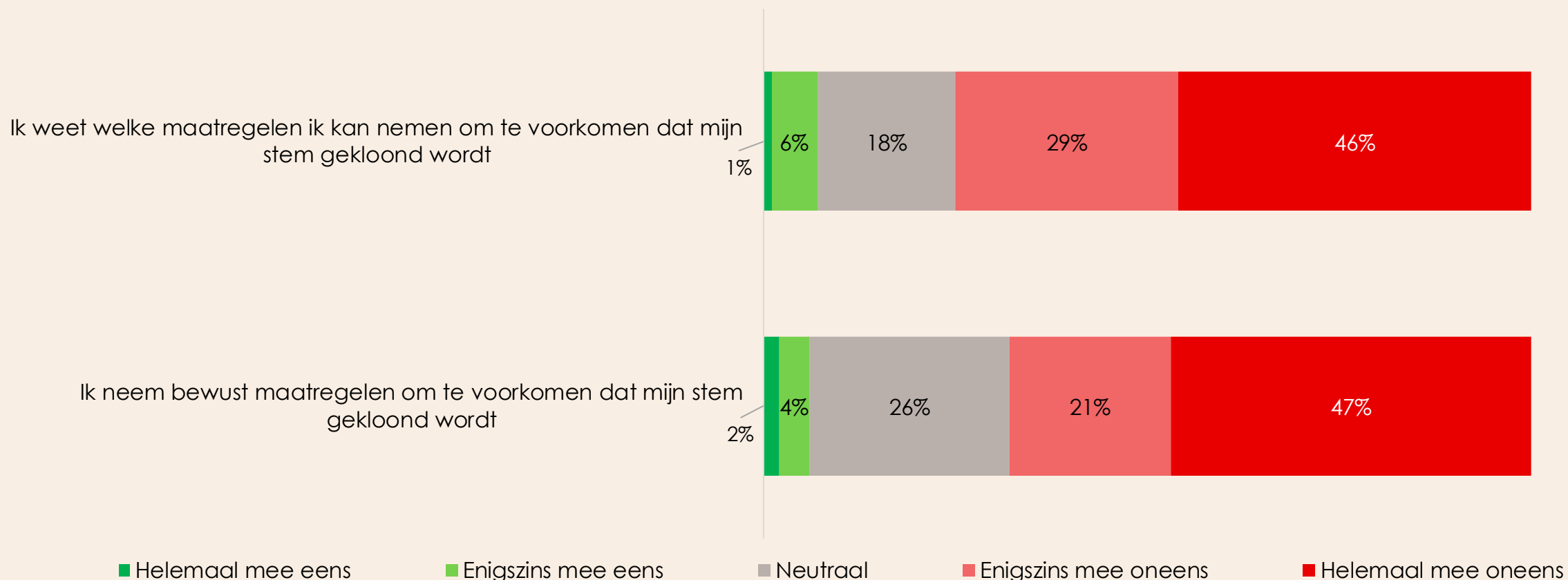
Bewustzijn en risicoperceptie stemklonen

Slechts 2% van de Nederlanders vindt het niet lastig om te onderscheiden of een stem echt of gekloond is. Ook denkt slechts 2% van de Nederlanders dat stemklonen in de toekomst niet zal worden gebruikt om mensen te bedriegen.



Beschermende maatregelen stemklonen

Er is onder Nederlanders nog veel onwetendheid over beschermende maatregelen: slechts 7% weet (enigszins) welke maatregelen tegen stemklonen beschermen, 75% weet het niet.

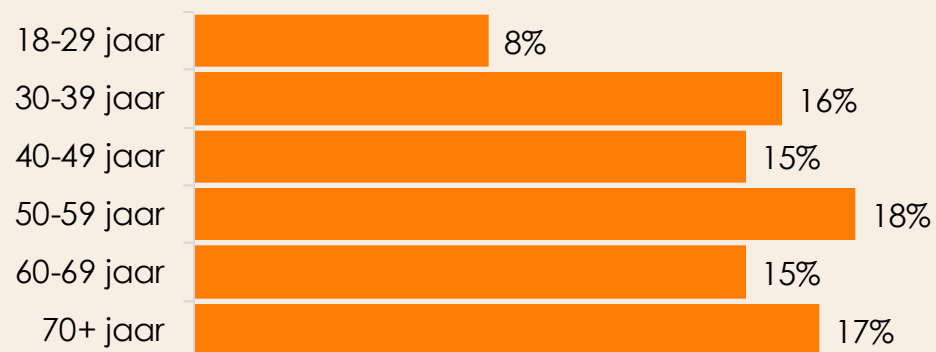


03

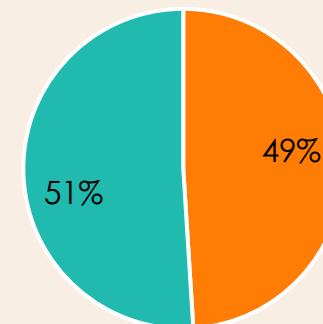
Achtergrondkenmerken doelgroep

Achtergrondkenmerken doelgroep

Leeftijd

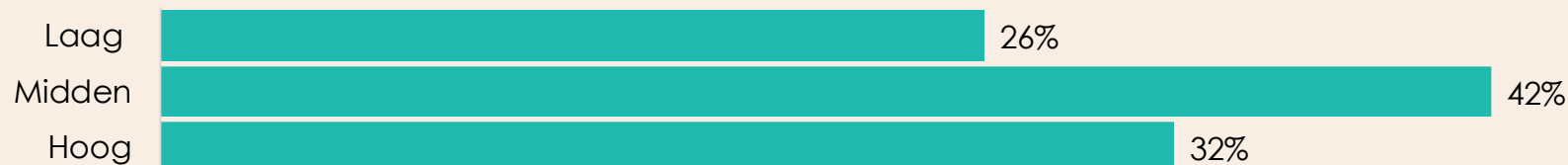


Geslacht



Man Vrouw

Opleiding*



04

Bijlage

Geluidsfragmenten – gekloond door stemtechnologie AI

Fragment 1:

Door de stromende regen moeten rennen om een trein te halen, en die dan net voor je neus weg zien rijden. Dat kan je dag echt verzieken.

Fragment 2:

De trein kwam precies op tijd, en ik had een zit plek. De reis ging helemaal volgens plan, en ik had alle tijd om naar een podcast te luisteren.

Fragment 3:

Op het nippertje een trein halen, terwijl je er eigenlijk al vanuit was gegaan dat je hem ging missen. Dat kan echt je dag maken.

Geluidsfragmenten – positief en negatief (BN-er en gekloond)

Positieve stemfragment

Je komt aanrijden en net als iemand vertrekt op die overvolle parkeerplaats; BAM! Het beste plekje precies voor de deur. Dat voelt als een geluksmomentje.

Negatieve stemfragment

Te laat door de files en dan nog 10 minuten ronddolen naar die parkeerplek, ver weg van je bestemming. Tsja, dat is gewoon balen.