



🕒 22 december 2020, 10:00 (CET)

# Argument-checking AI tegen fake news en misinformatie

Argumententechnologie-experts van de Universiteit van Amsterdam en MMGA bundelen hun krachten

De ISOC Nederland Make Media Great Working Group (MMGA) heeft besloten om samen te werken met twee leden van het KRINO-team, gericht op de ontwikkeling van een AI-aangedreven *argument-checking* tool. De samenwerking zal de gezamenlijke missie versterken om de trend van een toenemende hoeveelheid desinformatie en *fake news* die de wereld steeds meer verdeelt en elke dag meer impact heeft op de maatschappij, om te keren.

*Fact-checking* is tegenwoordig een belangrijke journalistieke handeling, maar omvat niet alle aspecten van het *debunken* van misinformatie en *fake news*: het verifiëren van de argumenten zelf lijkt een noodzakelijk en fundamenteel onderdeel van het proces van informatiekwaliteitscontrole. Dat is de kloof die KRINO vastbesloten is te overbruggen.

KRINO-teamleden, Dr. Federica Russo en Dr. Jean Wagemans, beiden werkzaam aan de Universiteit van Amsterdam, proberen nieuwe manieren te vinden om mensen hierover te onderwijzen, en hen in staat te stellen kritisch te denken. Kritisch denken is de analyse van de feiten die gebruikt worden om een oordeel te vormen. De vroegste gegevens over kritisch denken zijn de leerstellingen van Socrates die door Plato zijn vastgelegd. Nu in 2020 heeft Dr. Wagemans een gemoderniseerd kritisch denkveld gecreëerd, dat door het MMGA-team is omgevormd tot een hulpmiddel voor burgers buiten academische sferen om argumentatieve misvattingen te herkennen die in de media worden uitgebuit en die leiden tot verkeerde informatie, bijvoorbeeld in de vorm van onbetwiste politieke toespraken.

Uiteindelijk zal het informatievervuilingschild dat MMGA en KRINO aan het ontwikkelen zijn mensen een systematische filosofische bewaking bieden, voornamelijk voor hun online nieuwsconsumptie en stemoriëntatie, maar nuttig voor alle informatieve aspecten van het moderne leven.

De urgentie voor het werken aan deze zaken is dat de verspreiding van desinformatie en *fake news* de wereld steeds meer verdeelt en elke dag meer impact heeft op de maatschappij. De Amerikaanse verkiezingen hebben de dreiging van de zaak al bewezen en de Nederlandse verkiezingen zijn in volle gang. Er moet nu actie worden genomen. Het KRINO-team en MMGA hebben besloten samen te werken om elkaars missie te versterken en een sterkere en veerkrachtige gemeenschap te vormen om deze trend te keren.

KRINO wordt gezien als een transparante, verklaarbare kunstmatige intelligentie (XAI) die de gebruikers helpt bij het evalueren van argumenten. De argumenten kunnen in verschillende contexten voorkomen - van politieke manifestaties en toespraken, tot aanbevelingen van artsen, tot *social media posts*. In tegenstelling tot veel AI-gebaseerde tools zal KRINO geen compleet geautomatiseerde tool zijn, omdat dat het doel zou tarten om als menselijke hulp te dienen in plaats van de resultaten te dicteren uitsluitend op basis van algoritmen.

De focus van KRINO ligt op het ontwikkelen van een human-first AI-tool, inclusief de broodnodige ethische implicaties die voortdurend op de voorgrond worden geplaatst. Een deel van die filosofie is het creëren van een verklaarbare tool, niet een black box zoals veel AI-tools die momenteel worden geproduceerd. KRINO XAI is in volle ontwikkeling en zal grotendeels gebaseerd zijn op de lessen die het KRINO team zal verzamelen door samen te werken met de MMGA en ISOC gemeenschap.

Momenteel richten KRINO en MMGA zich volledig op het ontwikkelen van een gebruiksvriendelijk trainingstraject voor MMGA-annotators, ISOC-leden en het bredere online publiek om hun online weerbaarheid tegen informatievervuiling en retorietrucs tijdens de Nederlandse verkiezingen van 17 maart 2021 te vergroten. Als aftrap van de *argument-checking* training van KRINO-MMGA wordt begin 2021 een interactieve webinar gelanceerd.

Blijf op de hoogte van de laatste ontwikkelingen rond de komende MMGA argument-checking trainingen en KRINO als AI-tool!

Schrijf je in voor meer updates op de [MMGA-mailinglijst](#).

OVER KRINO

Krino AI is een tool om argumenten in geschreven natuurlijke taal te evalueren. Het is geen volledig geautomatiseerde tool, maar eerder een hulpmiddel voor de mens bij het analyseren van geschreven tekst en het ontwarren van kritische aspecten van de onderliggende argumentatiestructuur. In het bijzonder zal Krino helpen bij de verificatie van de informatie die in een tekst aanwezig is, in termen van correctheid en volledigheid, betaalbaar en toegankelijk voor elke competente gebruiker, wat resulteert in de onderbouwde overtuiging over geanalyseerde argumenten en besluitvorming. Krino wil een instrument zijn dat de gebruiker de mogelijkheid biedt om zijn cognitieve en sociale omgeving (en de informatie die hij analyseert) te verbeteren door deze transparanter, rationeler en begrijpelijker te maken. In die zin draagt Krino bij aan het bestrijden van misinformatie en het produceren van betere content. Krino is ontworpen als een inspecteerbare, white-box AI-engine die met de gebruiker communiceert in natuurlijke taal.

#### OVER ISOC NL MMGA WG

De Internet Society Nederland en Make Media Great Again hebben hun krachten gebundeld als de ISOC NL MMGA Working Group (WG). Samen met andere divisies van de Internet Society wereldwijd (de zogenaamde chapters) zet de ISOC MMGA Working Group een Global (mis)Information Observatory (GIO) op als een van de eerste prioriteiten. Meerdere ISOC-hoofdstukken binnen landen als Benin, India, Malta, Nederland, Slovenië en Suriname zijn al een samenwerking aangegaan met MMGA op het gebied van *fake news* webinars en gelokaliseerde verbetering van online nieuwsmedia. De werkgroep zal samenwerken met impactvolle media en wetenschappers, maar ook met de MMGA-gemeenschap bestaande uit kritisch denkende burgers en experts die "MMGA-annotators" worden genoemd. MMGA-annotaties zijn bruikbare suggesties in de vorm van gelabelde notities, die direct worden toegeschreven aan woorden, zinnen of paragrafen. Ze zijn bruikbaar voor de nieuwsredacteur, vermijden een debat op basis van persoonlijke voorkeuren en leiden, indien correct, direct tot een correctie binnen de artikelen. Het staat de redactie vrij om deze al dan niet te implementeren. Omdat de aantekeningen direct uitvoerbaar zijn en gebaseerd zijn op het principe van journalistieke objectiviteit, overwinnen ze het bekende probleem van het lange debat als gevolg van de subjectiviteit die ontstaat bij regelmatige lezerscommentaren.

---

### **What we do: Collaborating with publishers and community to fight misinformation**

We improve the quality of media together with their pool of knowledgeable readers, viewers and listeners. We have built a transparent system for actionable suggestions and specific remarks from our community pool of experts and critical thinkers.

NU.nl and AD.nl, two of the “Big Four” largest Dutch online news platforms, are our co-development partner and client. The system of MMGA is continuously tested with a screened and trained group of critical readers (‘annotators’) who offer constructive suggestions to increase the journalistic quality via balanced use of sources and clearer transfer of information.

### **Why we do it: improve the credibility of media and strengthen the bond with their audience**

The credibility of the media is being questioned more and more, whereas the media are the first parties to protect us from mis- and disinformation. This fundamental role of media is essential to enable proper functioning of democracy and constructive social debate, thus fortify social cohesion.

### **How we do it: Improving the quality of media through annotations**

We believe people have unique, diverse views and also relevant knowledge that helps the editorial process and quality.

With our digital tools people are able to detect misinformation, biased language and false contextualisation. MMGA introduces the notion of “annotations” in online journalism, initially in The Netherlands and subsequently worldwide. MMGA annotations are suggestions directly attributed to sentences or paragraphs. They are actionable for the editor, avoid debate based on personal preferences and, if correct, directly trigger a correction within articles.

