

Classificatiestandaard voor profielfoto's op social media

In onze onderzoeken hebben we vaak te maken met het bepalen van de authenticiteit van profielfoto's op social media platforms. We hebben hier een aantal geautomatiseerde tools voor en maken ook gebruik van visuele handmatige nacontrole. Bij beide methoden is het van belang dat de tools en de verschillende analisten eenzelfde standaard hanteren voor het indelen van de verschillende resultaten.

Hiertoe hebben wij een classificatiestandaard ontwikkeld voor het indelen van de resultaten van de geautomatiseerde en handmatige visuele controles.

Deze classificatiestandaard onderscheidt twee belangrijke criteria:

1. was de profielfoto bedoeld om de eigenaar van het account af te beelden?
2. indien ja: is de profielfoto ook daadwerkelijk een afbeelding van de eigenaar van het account?

Binnen deze onderverdeling zijn er nog verschillende subcategorieën te onderscheiden.

1 De foto moet de eigenaar van het account voorstellen

1.1 De foto is daadwerkelijk de eigenaar van het account

1.1.1 Accurate foto ("Accurate")

Dit is de meest voorkomende categorie: de profielfoto is daadwerkelijk een herkenbare en gelijkende foto van de eigenaar van het account. Ook komen foto's voor van de accounteigenaar samen met gezinsleden.

1.1.2 Tekening of avatar ("Impression")

De profielfoto is een tekening, tot tekening gefotoshopte foto of avatar, maar is wel herkenbaar als en gelijkend op de eigenaar van het account.

1.1.3 Onherkenbare afbeelding ("Disguised")

De profielfoto stelt daadwerkelijk de eigenaar van het account voor, maar deze is niet goed herkenbaar. Dat kan zijn doordat de foto vanuit een positie is genomen waarbij de persoon niet goed zichtbaar is, doordat een object (zoals een fototoestel die voor het gezicht wordt gehouden) een deel van het gezicht afschermt, of doordat de persoon is vermomd, verkleed of gemaskerd. Ook foto's met mondkapjes en sluiers vallen soms in deze categorie, eveneens als foto's van de betreffende persoon uit diens kindertijd.

Ook verknipte foto's waarbij er alleen details van het gezicht in close-up zichtbaar zijn (b.v. de ogen) behoren tot deze groep.

1.2 Degene op de foto is niet de eigenaar van het account

Dit is de categorie waar er sprake is van misleiding of inauthentiek gedrag. Ook hier kunnen we een aantal subcategorieën onderscheiden.

1.2.1 Gestolen foto (“Imposter”)

Het betreft hier een foto die van een ander account is gestolen waarbij de intentie is dat degene op de foto niet bij name wordt herkend. Het doel is niet zich voor te doen als de persoon op de foto, maar puur om een afbeelding van het account te verkrijgen. Doelstelling van de eigenaar van het account is niet herkenbaar te zijn maar wel een geloofwaardig profiel te hebben. Vaak wordt de foto verknipt om ‘omgekeerd zoeken’ te hinderen.

1.2.2 Identiteitsdiefstal (“Identity thief”)

Hier betreft het het gebruik van een foto waarbij het wél de bedoeling is zich als die persoon voor te doen. De foto maakt daarbij dus onderdeel uit van een “fake account” waarbij niet alleen de foto, maar alle gedragingen van het account zich voor proberen te doen als die betreffende persoon. Veel platforms hebben hier een “verified account” bedacht als tegenmaatregel. Dit gebeurt echter niet alleen bij beroemdheden, maar ook bij mensen die bekend zijn in een bepaalde niche of binnen een organisatie. Vaak zitten hier frauduleuze doelstellingen achter.

1.2.3 Gekaapt profiel (“Stolen profile”)

Dit is de omgekeerde situatie als bij **Identity Thief**. Een account van een persoon is gestolen en voorzien van een andere profielfoto en boodschap. Doel is vaak misbruik te maken van het netwerk van het oorspronkelijke profiel.

1.2.4 Gegeneerde foto (“Generated”)

In dit geval is de foto een door een neuraal netwerk gegeneerde profielfoto. Deze foto’s beschikken vaak over bepaalde technische en uiterlijke kenmerken waardoor ze als zodanig te herkennen zijn. Doelstelling van de eigenaar van het account is niet herkenbaar te zijn maar wel een geloofwaardig profiel te hebben. De techniek gaat snel vooruit en het wordt steeds lastiger het verschil te maken tussen **Generated**, **Imposter** en **Accurate**.

2 De foto moet niet de eigenaar van het account voorstellen

Er zijn ook veel mensen die hun profielfoto niet gebruiken om zichzelf af te beelden en waarbij dat ook helemaal niet de intentie is. Ook hier onderscheiden we vier categorieën.

2.1 Persoonlijke afbeelding (“Proud owner / parent”)

Afbeeldingen van iemand zijn auto, huisdier, fiets, kinderen, vakantiehuis, gebeurtenis, waarbij er een sterke, persoonlijke, affectie of relatie bestaat tussen de eigenaar van het account en de afbeelding. Het betreft hier mensen die trots zijn op iets en datgene als profielfoto gebruiken. De foto’s zijn in de regel zelfgemaakt.

2.2 Associatie (“Fanpic”)

Dit gaat over mensen die zich sterk associëren met een bepaalde organisatie of imago. Denk aan het gebruik van het logo van een voetbalclub, automerk of politieke partij. Ook stockphoto’s van bloemen en steden vallen in deze categorie.

Een speciaal geval zijn nog foto's van een acteur of actrice, meestal in een rol in een film. Hierbij is bijvoorbeeld te denken aan het gebruik van foto's van John Travolta in Pulp Fiction. Het doel is niet zich voor te doen als John Travolta of Vincent Vega, maar om de boodschap "Pulp Fiction is een coole film" over te brengen. De grens tussen **Fanpic** en **Identity Thief** zit hem vooral in de intentie.

2.3 Meme of politieke slogan ("Statement")

Hierbij gaat het om - vaak tijdelijke - profielfoto's die *en masse* worden vervangen in het kader van een hype. Doelstelling is het maken van een politiek statement: alleen de tekst #BLM of een volledige zwarte afbeelding zijn hier goede voorbeelden van. Ook het massaal plaatsen van een hamster als profielfoto op TikTok ("Hamster Cult") valt in deze categorie.

2.4 Geen profielfoto ("Blank")

Dit betreft de categorie die geen profielfoto heeft ingesteld en waarbij het account nog op de standaard foto staat, meestal een silhouet.