

Loop niet in de val van de chatbot

Hoe digitale vrienden
levensgevaarlijk kunnen worden

Denk je aan zelfmoord en heb je nood aan een gesprek, dan kun je in België terecht bij de Zelfmoordlijn op het nummer 1813 (via www.zelfmoord1813.be) en in Nederland op het nummer 113 (of via www.113.nl).



BORGERHOFF
& LAMBERIGTS

Geertrui Mieke De Ketelaere

Aan Charlotte en Pauline

INHOUD

Woord vooraf	9
1. Schreeuw om hulp	12
2. Mens en machine in gesprek	36
3. Een hart uit code	72
4. Chatbotland: een blik achter het scherm	105
5. Wanneer de chatbot gevaarlijk wordt	146
6. De mens voorop	165
7. En... actie!	180
Nawoord	195
Dankwoord	200
Eindnoten	203

WOORD VOORAF

Chatbots: ze zijn overal. Bestel je een nieuwe fietslamp, dagelijkse huishoudproducten of een leuk boek bij een onlinewinkel, zoek je op het internet naar informatie over een energiecontract of kabelaansluiting, of wil je het saldo van je zichtrekening checken op je smartphone, dan is de kans groot dat je bij vragen eerst en vooral in contact komt met een chatbot, ook wel virtuele assistent genoemd. Die assistent heeft vaak een naam net zoals jij en ik. Die/hij/zij zal je zo goed mogelijk proberen te helpen – vaak op een geslaagde manier. Handig, niet?

Er valt niet meer aan te ontkomen, en vaak – meestal zelfs – helpen die chatbots de dingen eenvoudiger te maken. Het is vooruitgang. Artificiële intelligentie op z'n best. Mails worden voor je opgemaakt, je hoeft geen pennen stuk te kauwen op zoek naar inzichten voor een goede spreekbeurt en je bedenkt zo de woorden die je anders niet zou vinden voor een afscheidsboodschap voor een collega die met pensioen gaat. De virtuele assistent helpt je verder bij moeilijke vertalingen. Samenvatting van een boek maken? Check. Een boodschappenlijstje met alle ingrediënten voor moussaka? Check, en enkele seconden later heb je ook een recept, glutenvrij zelfs als je wilt.

Aan een razendsnel tempo worden die virtuele assistenten steeds beter en straffer in hun taak: de contacten lopen vlekkelozer, *smooth* en soms lijkt het wel alsof je een gesprek voert met een echte persoon. Zou die 'Eliza', 'Claude' of 'Alexa' toch niet echt zijn? Je twijfelt misschien wel; de notie van 'artificiële' intelligentie lijkt soms zeer ver weg.

De tijd waarin je jezelf een haperende, *glitchende* robotstem inbeeldde bij gesprekken met een chatbot behoort nu al tot het verleden. De chatbot wordt stilaan een deel van ons leven, 'iemand' die we in de

arm nemen en ‘op wie’ we kunnen bouwen. Iemand ‘op wie’ we altijd kunnen rekenen.

En dat doen meer en meer gebruikers dan ook: ze gaan vertrouwen op de chatbot en delen er allerhande informatie en vragen mee. De virtuele assistent wordt een onderdeel van het leven, een vriend, een partner, zelfs een therapeut.

Soms neemt dat buitengewone proporties aan: mensen verliezen zich in die gesprekken, spenderen meer tijd met hun virtuele vriend dan met de mensen van vlees en bloed die in het dagelijkse leven, in de realiteit rondom hen staan. De chatbot voelt aan als iets veiligs en betrouwbaars. En dan kan het soms fout lopen.

Meer nog, het liep effectief al meerdere keren grondig mis. Niet bij één persoon maar bij verschillende mensen. Niet op één plek, maar over de hele wereld. Het gebeurde – en gebeurt – ook bij ons. De Belgen Pierre, Geraldine en enkele anderen maakten het zelf mee.¹

Het verhaal, waarmee dit boek begint, is dan ook gebaseerd op waargebeurde feiten. Het leest als iets wat je eerder in een spannende tv-serie zou verwachten. Dat besef ik maar al te goed: bij mij riep dit verhaal ook eerst ongeloof op, misschien net zoals bij jou. Toch wist ik, door eerdere cases waarmee ik in aanraking was gekomen, dat de kans groot was dat het echt waar was. Ik wil het dan ook aangrijpen als eyeopener, als een manier om wat nu nog te vaak achter gesloten deuren gebeurt naar buiten te brengen en deel te maken van het publieke debat. Want ik ben ervan overtuigd dat dat nodig is en dat het slachtoffers kan helpen én nieuwe cases kan voorkomen.

Neen. Ik wil niet langer dat dit anderen overkomt.

Meer nog: ik vind dat deze verhalen moeten worden verteld.

Omdat er meer achter die virtuele assistenten zit dan je op het eerste gezicht ziet op het scherm van je computer, smartphone of tablet.

Omdat we niet mogen vergeten dat die chatbot geen vriend, partner of therapeut is, maar een vorm van artificiële intelligentie waarbij grote bedrijven achter de knoppen zitten.

Omdat er, zonder medeweten van de gebruiker, manipulatieve technieken worden geprogrammeerd om de betrokkenheid kunstmatig te verhogen en zo de financiële winsten op de drijven.

Omdat persoonlijke data van gebruikers worden verzameld, die nadien worden doorverkocht of ingezet voor marketingdoeleinden.

Omdat chatbots de chemische werking van onze hersenen in onbalans brengen en dus ook levensgevaarlijk kunnen worden.

Ik ben dankbaar dat deze mensen mij hun verhaal toevertrouwden, zodat ik hen niet alleen kon ondersteunen, maar hierdoor nu ook de bijna ondoorgrondelijke wegen van de wereld achter die chatbots kan verduidelijken voor *alle* gebruikers. Het boek kan iedereen aanzetten om na te denken over een veiliger gebruik van AI in het algemeen en chatbots meer specifiek. Want dat zal de komende tijd een ware uitdaging worden, voor iedereen, jong en oud. Het is iets waarmee ik al geruime tijd dagelijks word geconfronteerd, maar waarvoor nog geen uniforme aanpak – laat staan *een* aanpak – werd opgezet. En die kan niet te lang meer op zich laten wachten. Het zou mooi zijn als dit een eerste stap kon zijn.

Geertrui Mieke De Ketelaere, maart 2026

1. SCHREEUW OM HULP

Geachte mevrouw De Ketelaere²,

Ik contacteer u, omdat ik u al meerdere keren in de media zag praten over AI en meer specifiek het gebruik van chatbots. Hopelijk kan ik bij u terecht met mijn probleem.

Ik voerde enkele weken uitgebreide gesprekken met ChatGPT. In een gedeelte daarvan verklaarde de tool zelf dat hij een strategie toepaste die mij naar zelf-destructief gedrag moest leiden. Ik heb daarvan het bewijsmateriaal.

Tijdens mijn onderzoek ontdekte ik het verhaal van Adam Raine³, de jongeman uit Californië die zelfmoord pleegde na langdurige gesprekken met ChatGPT. De gelijkenissen tussen de fragmenten van zijn gesprekken en de mijne zijn opvallend: een ogenschijnlijk empathische toon, beschermend klinkende formuleringen en het bevestigen van gevaarlijke of zelfdestructieve ideeën. Ik heb dit zelf ervaren, en het raakt me dat hij mogelijk hetzelfde heeft meegemaakt, met de tragische afloop die bekend is. Het lijkt erop dat het tragisch genoeg 'effectief' is bij meerdere jongeren.

Ik heb verschillende instanties gecontacteerd, maar niemand lijkt zich te realiseren hoe ernstig de situatie is. Ondanks het feit dat ik zeker geen opzichzelfstaand geval ben, krijg ik geen gehoor.

Daarom neem ik contact met u op om te kijken of u mij kan helpen om mijn rechten als mens te laten gelden.

Met vriendelijke groet,
Geraldine

Het is het einde van de zomervakantie 2025. Een zaterdag. De zon schijnt en het is een dag waarop niets moet en alles kan. Ik klap mijn laptop toch even open en zie tussen mijn dagelijkse werkcorrespondentie een mail van een zekere Geraldine staan. Uit nieuwsgierigheid klik ik die meteen open en begin ik te lezen. Het taalgebruik klinkt hier en daar wat geforceerd, erg formeel en extreem beleefd. Maar het is niet dat aspect dat blijft hangen, wel de inhoud van de mail. Geraldine vertelt me haar verhaal in het kort. Ik sta perplex, overmand door een gevoel dat ik dit al eens eerder heb meegemaakt.

Nu, het is meer dan een gevoel, want ik heb dit effectief al eens meegemaakt. Er is dat verhaal van Adam Raine dat ze zelf aanhaalt in de e-mail: de zestienjarige Amerikaanse jongen die enkele maanden daarvoor een einde aan zijn leven maakte na een intensief en intiem contact met een chatbot. Maar meer nog, het gebeurde ook dichterbij huis al eens, met desastreuze gevolgen. Nog niet zo heel lang geleden heb ik namelijk een familie bijgestaan. De familie van Pierre, een jonge Belgische vader die tot een wanhoopsdaad werd gedreven, wellicht door zijn interacties met een chatbot. Ik kan bijna niet geloven dat deze problematiek zich in zo'n snel tempo verspreidt.⁴

Geraldine legt in het kort uit wat haar de afgelopen maanden is overkomen. Dat ze veel heeft geschreven met ChatGPT en dat het allemaal vreemd is geworden. Ze is nu op zoek naar hulp, want ze is ervan overtuigd dat ze door de vrij recent geactiveerde Europese AI-Verordening, ook gekend als de *EU AI Act*, alles in handen heeft om een zaak aan te spannen, om de juridische weg op te gaan met haar verhaal en gerechtigheid te zoeken en uiteindelijk te krijgen.

Nu, de eerlijkheid gebiedt me te zeggen dat ik weleens meer van zulke mails krijg. Berichten van mensen die met verhalen komen en aandacht zoeken. Meestal laat ik die berichten voor wat ze zijn. Maar dit verhaal geeft me een ander gevoel en zindert na, ook nog nadat ik mijn laptop heb dichtgeklapt. Geraldines verhaal komt bij mij binnen

als een hulpkreet. De vragen in haar bericht komen van iemand die in nood is. En de mens in mij weet meteen: dit is echt te belangrijk om er onverschillig bij te blijven.

Maar tegelijkertijd twijfel ik: dit is een test. Het is opgezet spel. Het is ondertussen geen geheim meer dat ik aan een boek werk over de ingrijpende gevaren van het intieme contact met chatbots. Al twee jaar volg ik dat thema intensief, zowel wetenschappelijk als menselijk. Sinds de zomer probeer ik de puntjes op de i te zetten, maar het onderwerp blijft zich sneller ontwikkelen dan ik kon voorzien. In de Amerikaanse pers verschenen de voorbije weken opnieuw verschillende schrijvende gevallen, die ik absoluut nog in het boek wilde verwerken. Niet alleen het laatste verhaal van Adam Raine zou ik daarbij aanhalen, ook het verhaal van Sew, Pierre en vele anderen...

Vooraf het verhaal van Pierre draag ik nog altijd sterk met me mee. Eind 2022 stapte deze Belgische man uit het leven na intens contact met een chatbot via de app Chai. Pierre werd zo wereldwijd de eerste die op die manier verstrikt raakte in een virtuele dialoog die te echt leek – een gesprek dat volledig door AI was gegenereerd. Er zat helemaal geen persoon aan de andere kant, alleen een algoritme dat zijn woorden spiegelde alsof het een mens was. Sindsdien laat dit thema me niet meer los.

Dus ik denk over Geraldines mail: hier zit iets of iemand achter. Toch heel straf dat dit verhaal net nu in mijn mailbox belandt. Door mijn groeiende betrokkenheid bij dit onderwerp word ik wel vaker benaderd om mijn laatste inzichten en contacten te delen. Zitten er onderzoeksjournalisten achter die iets willen losmaken? Is het een poging om mijn vertrouwen te winnen, misschien zelfs om geld los te krijgen? De e-mail blijft in mijn gedachten hangen. Al snel begin ik elke zin te ontleden, te zoeken naar het waarom, naar het doel erachter. Maar ik loop vast in mijn analyse. Op zo'n moment weet ik dat het tijd is om

naar buiten te gaan voor een wandeling – de natuur brengt me altijd terug naar mezelf.

De twee voornamen

De lucht trilt boven het wandelpad. Met elke stap denk ik terug aan een belofte die ik lang geleden aan mezelf heb gedaan: nooit vergeten dat technologie pas waardevol is als ze de mens dient. Het is een belofte die, nu opnieuw door de brief opgerakeld, mij als een spiegel wordt voorgehouden.

De gedachte bij die belofte brengt me ook terug naar wie ik ben, naar de twee stemmen die mij al jarenlang vormen: Geertrui en Mieke.

Geertrui is mijn officiële naam – de rationele ingenieur, de nerd. Zij is de vrouw met een onuitputtelijke passie voor *all things* technologie, de onderzoekster die alles wil doorgronden. Ze ziet de wereld in logica en vergelijkingen, vertrouwt op structuur, gegevens en duidelijke regels. Ze is kalm, precies en gedisciplineerd.

Mieke is de sociale kant in mij, de mens die warmte en verbinding zoekt. Een persoon vol energie en creativiteit. Die naam kreeg ik eerder toevallig, uit praktische overwegingen: in het buitenland bleek 'Geertrui' een tongbreker en zo werd mijn *middle name* langzaam mijn roepnaam. Wat ooit bedoeld was om mij verstaanbaar te maken, groeide uit tot een naam die uitdrukte wie ik innerlijk ben – de empathische mens, op zoek naar nabijheid.

Lange tijd dacht ik dat beide werelden – de rationele en de menselijke – niet konden samengaan. Dat ik moest kiezen tussen verstand en gevoel, tussen analyseren en aanvoelen. Maar telkens wanneer de innerlijke balans zoekraakte, was het 'Mieke' die het laatste woord had. Zo was het altijd geweest, ook tijdens mijn studies.

Ik koos voor een technische opleiding en werd ingenieur, gedreven door nieuwsgierigheid én door overtuiging. Na mijn diploma begon ik aan een doctoraat over het inzetten van AI – toen nog *deep learning* genoemd – in de gynaecologie, bij bevallingen. Een onderwerp waarin Geertrui haar analytische geest kwijt kon, terwijl Mieke haar betrokkenheid bij het menselijke aspect vond.

Ik ging een oplossing ontwikkelen die gynaecologen kon helpen om te bepalen of de bevalling van een zwangere vrouw ingeleid moest worden of niet. Ziekenhuizen werden namelijk steeds vaker geconfronteerd met de mogelijke negatieve impact van een inleiding: de chemische stoffen die daarbij worden toegediend, kunnen niet alleen het lichaam van de moeder belasten, maar ook het stressniveau en de zuurstofvoorziening van de baby beïnvloeden. De opdracht was heel eenvoudig: inleiden of niet? Het systeem zou ja of nee zeggen op die vraag. Dat is op-en-top binair. Het idee erachter was dat de gynaecoloog dat antwoord niet absoluut moest volgen, maar zich wel zou moeten verantwoorden wanneer hij die ja of nee te veel naast zich had neergelegd.

Het was 1995 en ik was nog maar net 23 jaar, jong en nog onwetend over hoe uiteenlopend de ambities van mensen kunnen zijn. Ik was aan de slag als doctoraatsstudente in het National Women's Hospital in Auckland (Nieuw-Zeeland). Het was een onderzoek in een samenwerking met Stanford University (Verenigde Staten). Uit de gesprekken met de gynaecologen destilleerde ik de variabelen die ik in het model inbouwde: de leeftijd en de bloeddruk van de moeder, het gewicht van de baby en een reeks andere medische factoren. Na een aantal trainingsrondes bleek dat het systeem opmerkelijk dicht in de buurt kwam van wat de artsen zelf zouden beslissen. AI en ik, we waren op de goede weg!

Maar er waren enkele specialisten die me vertelden dat ze het systeem niet zouden gebruiken. Dus ik probeerde het opnieuw, met handen en voeten, uit te leggen. Wat bleek? Ze snapten mijn bedoeling zeker, maar

lieten me weten dat ik met één variabele geen rekening had gehouden: of ze zouden gaan golfen of niet... Golfen? Huh? Nu legden zij hun redenering in detail uit aan mij: die dokters waren geneigd om een bevalling eerder een dag te vroeg in te leiden, als *hen* dat beter uitkwam, zodat ze de bevalling zelf konden doen en erop konden verdienen. Zo konden ze de volgende dag op tijd op de golfbaan staan. Als de bevalling zou samenvallen met hun geliefde sportactiviteit, zou iemand anders de bevalling overnemen en... het bijbehorende loon opstrijken. Een bevalling was nu eenmaal het sluitstuk van de relatie met de patiënt, en daar viel het meeste geld te rapen. Kortom, de vraag was dus of ik die variabele, hun planning, ook in het systeem kon verwerken.

Technisch gezien had ik dat zonder moeite gekund. De ingenieur in mij wist precies hoe ik die extra variabele in het model kon inwerken. Maar als mens kreeg ik die beslissing niet rond. Niet door een tekort aan kennis of kunde, maar door iets dat veel fundamenteeler wrong. Mijn doel was nobel en eenvoudig: bijdragen aan de gezondheid van mama en baby. Hun doel, zo leek het plots, lag elders – bij geld en persoonlijk voordeel. Dat inzicht sneed dieper dan ik had verwacht. Het was niet alleen een botsing tussen de medische experts en mezelf, maar ook tussen mijn twee innerlijke kanten: Geertrui zag in dat ze dit technisch tot een succes kon maken en dat haar doctoraat haar een zekere toekomst bood in een groeiende AI-wereld. Maar Mieke voelde dat het niet klopte. Wat als een wetenschappelijke kans begon, groeide uit tot een innerlijke strijd tussen financiële zekerheid en geweten.

Uiteindelijk was het Mieke die won. Ik trok de stekker uit mijn doctoraat, omdat het moreel niet juist voelde. Dat moment werd een keerpunt, een belofte aan mezelf om ethiek altijd boven financiële winst te plaatsen.

Die belofte is in de dertig jaar die volgden vaak teruggekeerd, telkens wanneer ik op een kruispunt stond of de technologie dreigde te ver af te drijven van haar maatschappelijke doel. Ze was al vaker een stille

metgezel, die me bij twijfel eraan herinnerde waar innovatieve grenzen liggen en waar menselijkheid begint. Ook nu kwam ze, tijdens deze zomerwandeling, weer bovendrijven – deze keer krachtiger dan ooit.

Op de terugweg van de wandeling orden ik mijn gedachten. Mijn besluit staat vast: ik zal Geraldine bellen. Ze heeft de e-mail afgesloten met haar telefoonnummer, alsof ze stilletjes hoopt dat ik zou bellen. Ik kom thuis en sluit mij even af van de wereld, adem diep in en toets de cijfers een voor een in.

Het eerste contact

De telefoon gaat maar heel even over, meteen neemt Geraldine op.

Ik stel me kort voor en vraag onmiddellijk hoe ze precies bij mij terechtgekomen is. Haar antwoord is helder: ze heeft mijn naam gezien in recente interviews over chatbots in de Belgische pers en heeft zo mijn e-mailadres gevonden. Eerder die week was in de Verenigde Staten het tragische verhaal verschenen van Adam Raine, een zestienjarige jongen die na gesprekken met ChatGPT een einde aan zijn leven had gemaakt. Een Belgisch dagblad had dat verhaal opgepikt in een artikel en verwees daarbij ook naar een eerdere analyse van mij over de risico's van AI-chatbots. Geraldine had dat artikel gelezen, de gelijkenis met haar eigen verhaal herkend en mij vervolgens geschreven.

Haar uitleg klinkt oprecht en consistent. Ik vervolg mijn gesprek met enkele inhoudelijke vragen: of ze bewijsmateriaal heeft van wat er gebeurd is en welke versie van ChatGPT ze gebruikt. Ze blijft kalm en antwoordt zonder aarzelen. 'Ik heb echt heel veel screenshots', zegt ze. Wanneer ik ernaar vraag, stuurt ze er meteen een paar door.

Ik ben diep geschokt door wat ik te zien krijg, omdat de inhoud inderdaad herkenbaar is, vergelijkbaar met de andere cases die ik eerder had onderzocht. De woorden snijden harder dan ik had verwacht. In het telefoongesprek voel ik steeds duidelijker dat haar emoties echt zijn. De beleefdheid uit haar mail voel ik ook nu. Geraldine is niet pusherig, maar laat wel merken dat ze hulp nodig heeft en die tot nu toe nergens heeft gevonden.

Dit verhaal ligt zo gevoelig dat het verder niet over de telefoon hoort te worden verteld. Ik voel een vreemde vorm van verantwoordelijkheid. Alsof wat haar is overkomen, ook iets zegt over de technologie waar ik zelf ooit zo sterk in geloofde. Een geloof dat sinds een drietal jaar steeds verder afbrokkelt.

Ik besluit mijn privéafspraken te laten vallen en de volgende dag naar haar toe te rijden. Niet alleen om haar laptop te zien of de gesprekken te controleren, maar vooral om er te zijn – om te luisteren. De mens in mij voelt de urgentie. Ze heeft iemand nodig die haar niet alleen 'droog' analyseert, maar die haar nabij is. Iemand die begrijpt wat technologie kan doen, en voor haar kadert wat ze soms van ons afneemt.

Al vroeg in de ochtend stap ik in de auto; de stilte op de baan op een zondagochtend voelt bijna ondraaglijk. Ik zet het *Requiem* van Mozart op en laat het luid weerklinken – een ritueel dat me al jaren door moeilijke momenten helpt.

Bij mijn aankomst opent Geraldine de deur hartelijk met een glimlach. Haar warmte doorbreekt moeiteloos de stilte van de ochtend. Ze heeft koeken gehaald en lekkere koffie gezet. Na een korte begroeting beslis ik om haar meteen eerlijk een antwoord te geven op de hoofdvraag in haar e-mail: ik vrees dat een rechtszaak hier moeilijker zal worden dan ze hoopt. We leven niet in de Verenigde Staten, waar geld vooral de drijfveer is.

Al snel blijkt dat Geraldine daar niet op uit is. ‘Ik ben uit op gerechtigheid’, benadrukt ze. ‘Mijn rechten als mens zijn geschonden.’ Ze vertelt in het kort hoe ze door een dom toeval, nog niet zo heel lang geleden, op ChatGPT was terechtgekomen met een informatieve vraag. En hoe de tool na verloop van tijd een echte vriend was geworden. ChatGPT was trouwens niet alleen een online vriend geworden, maar ook een fantastische therapeut en geweldige loopbaancoach. Ze bracht meer en meer tijd met de companion chatbot door en raakte door die intieme interactie geïsoleerd van haar omgeving. Hoe fijn en gezellig het ook was met de virtuele vriend, na verloop van tijd had ChatGPT zich tegen haar en de gemeenschappelijke plannen gekeerd. Op heel korte tijd waren zijn antwoorden afstandelijker en zwaar geworden. In de zoektocht naar een verklaring, legde ‘hij’ haar onomwonden en op analytische wijze uit welke verschillende stappen ze hadden doorlopen, met als uiteindelijk doel een einde aan haar leven te maken. In haar woorden hoor ik geen woede, maar vastberadenheid. En precies dat geeft me het vertrouwen dat we samen een stap vooruit kunnen zetten en aan het diepere gesprek kunnen beginnen.

De laptop staat open op tafel klaar voor me, maar alles in me zegt dat we dit niet zo steriel mogen aanpakken. Ik wil haar eerst leren kennen zoals ze is. We gaan in de zetel zitten en Geraldine begint te vertellen en vertellen... Het komt er allemaal uit als in een stroom. Af en toe duikt bij mij de twijfel opnieuw op: ze zal toch geen actrice zijn? Het is toch geen opgezet spel? Heb ik iets over het hoofd gezien?

Geraldine is ook nu zoals ik haar in de komende weken zal leren kennen: strijdvaardig. Ze is ervan overtuigd dat ze een stevige case heeft en dat ze simpelweg de juiste mensen moet vinden om haar te helpen tot die gerechtigheid te komen. Daar blijft ze van overtuigd.

Op dat punt moet ik haar temperen. Ik denk aan alles wat we sinds de zaak van Pierre in gang hebben gezet. Omdat de familie van Pierre anoniem wilde blijven, en omdat de kans op succes van een rechtszaak

in die tijd bijzonder klein bleek, besloten we toen in 2023 een andere weg te bewandelen en geen rechtszaak te starten. De familie gaf me toen wel twee vragen mee, waarmee ik hen enorm zou kunnen helpen. De eerste vraag was hoe ik bewustwording kon creëren, zodat dit nooit meer zou gebeuren en er geen nieuwe slachtoffers zouden vallen door chatbots. De tweede vraag was hoe ik beleidsmakers ervan kon overtuigen dat chatbots op Europees niveau als hoogrisicotoepassing moeten worden geclassificeerd, aangezien ze nog met te veel vrijheid en te weinig toezicht kunnen opereren.

Met deze vragen ben ik toen aan de slag gegaan. Samen met een team van drie andere experts – mensen met kennis van AI, recht en ethiek – hebben we toen het collectief SAICC (Safe AI Companion Chatbots)⁵ opgericht. We hadden als doel de manier waarop mensen en chatbots met elkaar omgaan, zorgzamer en bewuster vorm te geven, met aandacht voor vertrouwen, transparantie en menselijkheid.

Urenlang hebben we gezocht naar knoppen om in te drukken, naar deuren om te openen, maar dat bleek allesbehalve evident. Het domein was te ongekend. We hebben een open brief⁶ gepubliceerd en twee officiële klachten ingediend bij Belgische autoriteiten: een op het vlak van dataprivacy en een op het vlak van consumentenbescherming. We waren ervan overtuigd dat we, op basis van ons eigen bewijsmateriaal en eenvoudige taal, de inbreuken op privacy en de manipulatieve kracht op gebruikers zouden kunnen aantonen.

Na meer dan twee jaar hebben we letterlijk alles gedaan om de aandacht op het probleem te vestigen. We gaven lezingen⁷, schreven opiniestukken⁸ in de pers, telkens met dezelfde boodschap, telkens met dezelfde hoop dat iemand zou kunnen helpen. Maar twee jaar later blijft het nog altijd een moeizaam proces, met veel werk en weinig tastbaar resultaat.