



BELEIDSPLAN 2025 – 2028

Stichting Aithos

www.aithos.org

28 juli 2025

Inhoudsopgave

1 Inleiding	2
2 Missie en Visie	3
2.1 Het Probleem van AI-Waardemonopolie.....	3
2.2 Pluraliteit als Antwoord.....	3
2.3 Van Privaat naar Publiek.....	3
2.4 Maatschappelijke Problemen die Aithos Aanpakt.....	4
2.5 Doelstellingen 2025-2028.....	5
2.6 Onze Aanpak.....	6
2.7 Evaluatiecriteria.....	6
3 Huidige Situatie en Activiteiten	7
3.1 Kernactiviteiten 2025-2028.....	7
3.2 Praktijkvoorbeeld: Zichtbaar Maken van AI-Waarden in Klantenservice...	9
4 Toekomstige Ontwikkelingen	10
4.1 Positionering Aithos.....	11
5 Organisatiestructuur	12
5.1 Bestuur.....	12
5.2 Uitvoerende Organisatie.....	12
5.3 Samenwerkingsnetwerk.....	12
6 Integriteitsprotocol	14
6.1 Conflictenbeleid (Overgangsfase).....	14
6.2 Ethische Principes voor Bestuur:.....	14
6.3 Klokkenluidersregeling:.....	14
7 Financiële Planning	16

1 Inleiding

AI-systemen nemen steeds vaker zelfstandig beslissingen die direct impact hebben op mensenlevens. Chatbots beantwoorden vragen over verzekeringen en uitkeringen, AI-systemen helpen bij het beoordelen van leningen, en geautomatiseerde systemen ondersteunen beslissingen in de zorg en het onderwijs. Deze ontwikkeling confronteert ons met een fundamenteel politiek probleem: wie bepaalt welke waarden deze systemen hanteren, en hoe kunnen verschillende perspectieven samenleven zonder dat één visie wordt opgelegd aan alle anderen?

Het probleem gaat dieper dan technische fouten. AI-systemen bevatten altijd impliciete oordelen over wat belangrijk is, hoe problemen moeten worden opgelost, en welke afwegingen acceptabel zijn. Deze oordelen worden meestal gemaakt door kleine groepen technici en bedrijfsleiders, zonder publieke discussie of democratische legitimatie. Het resultaat is wat we 'privatisering van waardenkeuzes' kunnen noemen – fundamentele beslissingen over hoe onze samenleving functioneert worden genomen achter gesloten deuren.

Aithos is als stichting opgericht om AI-waarden terug te brengen naar de publieke sfeer. Geïnspireerd door het begrip van pluraliteit – het idee dat gemeenschappen juist gedijen van de spanning tussen verschillende perspectieven – ontwikkelen wij methoden om AI-waardenkeuzes zichtbaar te maken, om verschillende visies naast elkaar te laten bestaan, en om ruimte te creëren voor publieke en private discussie over hoe we willen dat deze technologieën functioneren.

We geloven niet dat er één juiste set waarden bestaat die alle AI-systemen zouden moeten hanteren. In plaats daarvan werken we aan het creëren van mechanismen waarmee verschillende gemeenschappen, organisaties en individuen kunnen bepalen welke waarden hun AI-systemen moeten weerspiegelen, om zichzelf te kunnen vertegenwoordigen in een maatschappij waarin AI een centrale rol speelt.

In de periode 2025–2028 zal Aithos haar werk uitbreiden door samen te werken met bedrijven, kennisinstellingen en internationale maatschappelijke organisaties. Om deze reden benodigt Aithos ANBI-status, zodat we belastingvrije schenkingen kunnen ontvangen ter ondersteuning van onze democratiseringsmissie.

Dit beleidsplan beschrijft onze plannen voor de komende jaren, opgesteld conform de doelstellingen in onze statuten, met aandacht voor transparantie, goed bestuur en duidelijk maatschappelijk nut. Het wordt publiekelijk gedeeld op de website van Aithos.

2 Missie en Visie

2.1 Het Probleem van AI-Waardemonopolie

AI-systemen functioneren nooit neutraal. Elk algoritme bevat impliciete oordelen over wat belangrijk is, hoe afwegingen moeten worden gemaakt, en welke uitkomsten gewenst zijn. Een AI die ondersteunt bij het afnemen van sollicitaties moet bijvoorbeeld kiezen tussen het waarderen van creativiteit versus betrouwbaarheid, tussen het belonen van ervaring versus het geven van kansen aan nieuwkomers, tussen het prioriteren van technische vaardigheden versus sociale competenties.

Deze keuzes zijn inherent politiek – ze weerspiegelen visies op hoe de samenleving zou moeten functioneren. Het probleem is dat deze waardenkeuzes momenteel worden gemaakt door een kleine groep mensen (voornamelijk technische ontwikkelaars en bedrijfsleiders) zonder publieke discussie of democratische legitimatie. Er bestaan geen functionele en geaccepteerde evaluatiemethoden om inzicht te krijgen in deze waardesystemen, dus hoewel er door vele mensen ongewenst gedrag wordt opgemerkt, hebben gebruikers noch ontwikkelaars van AI momenteel een duidelijk overzicht. Gebrek aan inzicht in de waarden achter algoritmes heeft het afgelopen decennium al bijgedragen aan vele crises, van de toeslagenaffaire in Nederland tot de Rohingya-genocide in Myanmar.

Hannah Arendt, filosofe die uit Duitsland vluchtte voor het nationaalsocialisme en zich daarna verdiepte in totalitarisme, waarschuwde voor situaties waarin politieke kwesties uit de publieke sfeer verdwijnen en worden gereduceerd tot technische of administratieve problemen. Precies dit gebeurt met AI-waarden: fundamentele vragen over hoe we willen samenleven worden verscholen achter technische complexiteit en gepresenteerd als neutrale optimalisatieproblemen.

2.2 Pluraliteit als Antwoord

Arendt benadrukte dat politieke gemeenschappen hun vitaliteit ontleen aan de spanning tussen verschillende perspectieven – wat zij ‘pluraliteit’ noemde. Echte politiek ontstaat wanneer mensen met verschillende visies samenkomen om te overleggen, informeren en het conflict aangaan over gemeenschappelijke kwesties, niet wanneer één perspectief wordt opgelegd aan alle anderen.

Dit concept is cruciaal voor het begrijpen van AI-waarden. In plaats van te zoeken naar universele principes die alle AI-systemen zouden moeten volgen, moeten we ruimte creëren waarin verschillende waardesystemen kunnen samenleven. Een AI-systeem dat wordt gebruikt door een progressieve NGO hoeft niet dezelfde prioriteiten te hebben als één gebruikt door een conservatieve religieuze organisatie. Een systeem in de zorg kan andere afwegingen maken dan één in de financiële sector.

De uitdaging is niet om consensus te bereiken over welke waarden ‘juist’ zijn, maar om mechanismen te ontwikkelen waarmee verschillende waardesystemen transparant kunnen worden gemaakt, publiek besproken, en toegepast zonder anderen te domineren.

2.3 Van Privaat naar Publiek

Momenteel worden AI-waarden bepaald in private boardrooms en ontwikkelteams. Deze waarden worden vervolgens opgelegd aan miljoenen gebruikers zonder dat deze ooit de

kans hebben gehad om ze te bespreken, laat staan te beïnvloeden. Aithos werkt aan het terugbrengen van AI-waarden naar de publieke sfeer. Dit betekent:

Zichtbaar maken: AI-waardenkeuzes expliciet maken in plaats van ze verborgen houden in technische systemen.

Bespreekbaar maken: De waardesystemen van openbare modellen publiceren, zodat belanghebbenden zelf kunnen kiezen welk systeem ze het best vertegenwoordigt.

Beïnvloedbaar maken: Mechanismen ontwikkelen waarmee gemeenschappen, instanties, en individuen daadwerkelijk invloed kunnen uitoefenen op de AI-systemen waar ze mee te maken hebben.

Respecteren van verschillen: Zorgen dat het zichtbaar maken van waardenverschillen niet leidt tot het opleggen van één perspectief, maar tot het respecteren van legitieme diversiteit en keuzevrijheid.

Missie: Aithos brengt AI-waardenkeuzes terug naar de publieke sfeer door deze zichtbaar, bespreekbaar en beïnvloedbaar te maken, terwijl we ruimte creëren voor het vreedzaam samenleven van verschillende waardesystemen.

Visie: Aithos streeft naar een samenleving waarin:

- AI-waardenkeuzes onderwerp zijn van publieke discussie in plaats van private besluitvorming
- Technologie het publieke debat faciliteert zonder het over te nemen
- Burgers kunnen begrijpen en beïnvloeden welke waarden de AI-systemen hanteren die hun leven beïnvloeden
- Pluraliteit van waardesystemen wordt gerespecteerd in plaats van gereduceerd tot één dominante visie
- Verschillende gemeenschappen hun eigen AI-waarden kunnen bepalen zonder deze op te leggen aan anderen

2.4 Maatschappelijke Problemen die Aithos Aanpakt

We observeren de volgende trends in AI die tot grote maatschappelijke problemen zouden kunnen leiden. Ons onderzoek heeft als doelstelling deze problemen te voorkomen en verhelpen:

Depolitisering van waardenkeuzes: AI-waarden worden gepresenteerd als technische optimalisatieproblemen in plaats van als politieke keuzes die publieke discussie verdienen. Dit ontkoppelt fundamentele beslissingen over hoe onze samenleving functioneert van democratische controle.

Waardenmonopolie: Een klein aantal grote technologiebedrijven bepaalt de waarden die miljoenen AI-systemen hanteren, waardoor hun specifieke visies op maatschappelijke prioriteiten worden opgelegd aan diverse gemeenschappen met andere behoeften en overtuigingen.

Gebrek aan transparantie: AI-waardenkeuzes blijven verborgen achter technische complexiteit, waardoor het onmogelijk is om ze te begrijpen, laat staan te bespreken of aan te passen.

Pseudo-universalisme: AI-systemen doen alsof ze neutrale, universele oplossingen bieden, terwijl ze in werkelijkheid specifieke waardesystemen belichamen die niet geschikt zijn voor alle contexten en gemeenschappen.

Vervlakking van pluraliteit: De complexiteit en diversiteit van menselijke waardesystemen wordt gereduceerd tot simplistische optimalisatie doelen, waardoor de rijkdom van verschillende levensbeschouwingen en prioriteiten verloren gaat.

Ondemocratische besluitvorming: Beslissingen die grote impact hebben op hoe mensen leven en werken worden genomen zonder enige vorm van democratische

legitimatie of publieke betrokkenheid.

2.5 Doelstellingen 2025-2028

Om een oplossing te bieden voor de maatschappelijke problemen waar Aithos aan wil werken, zijn de volgende doelstellingen gekozen voor de komende drie jaar:

Zichtbaar maken van AI-waarden:

1. Ontwikkelen van tools en technieken die de verborgen waardenkeuzes in AI-systemen expliciet maken
2. Onderzoeken welke waardensystemen verschillende AI-systemen hanteren en hoe deze ontstaan
3. Creëren van begrijpelijke representaties van AI-waarden voor niet-technisch publiek

AI menselijke waarden laten uitdragen:

1. Ontwikkelen van methodes om AI-systemen de waarden van mensen of organisaties te laten vertegenwoordigen
2. Onderzoeken hoe AI-systemen consistent goed afgewogen morele beslissingen kunnen maken
3. Rollen waarin AI wordt ingezet en de bijbehorende behoeften en uitdagingen in kaart brengen

Implementeren van Pluraliteit:

1. Ontwikkelen van kaders waarin verschillende waardensystemen kunnen samenleven
2. Adviseren over governance-structuren die diversiteit accommoderen in plaats van elimineren
3. Ontwikkelen van methoden voor democratische deliberatie over technologische keuzes

Democratiseren van AI-Governance:

1. Ondersteunen van overheden bij het ontwikkelen van participatieve AI-beleid
2. Adviseren over toezichtsmechanismen die ruimte laten voor waardendiversiteit
3. Promoten van bottom-up benaderingen van AI-regulering

2.6 Onze Aanpak

Bij het realiseren van onze doelstellingen wordt de volgende strategie voor ogen gehouden:

Anti-Utopisch: Wij zoeken niet naar de ‘juiste’ waarden voor AI, maar naar methoden om verschillende legitieme waardensystemen zichtbaar te maken en naast elkaar te laten bestaan.

Pluraliteit-Gericht: In plaats van consensus na te streven, werken wij aan het creëren van ruimte waarin verschillende perspectieven beraadslaagd kunnen worden zonder dat één visie dominant wordt.

Publiek-Politiek: Wij brengen AI-waardenkeuzes uit de private sfeer naar de publieke arena waar ze kunnen worden besproken door alle betrokkenen.

Context-Afhankelijk: Verschillende situaties en gemeenschappen hebben verschillende behoeften. Onze methoden respecteren deze diversiteit in plaats van deze weg te reduceren.

Democratisch-Participatief: Wij ontwikkelen methoden waarmee gewone burgers en gemeenschappen daadwerkelijk invloed kunnen uitoefenen op AI-systemen die hen beïnvloeden.

2.7 Evaluatiecriteria

Om deze doelstellingen te realiseren streeft Aithos naar de volgende concrete kwaliteitsindicatoren in de komende drie jaar:

Doel	Indicator
Wetenschappelijke impact	• Publicaties in erkende tijdschriften (Q2 of hoger) • Deelname aan academische conferenties
Maatschappelijk gebruik	• Overheidsinstanties of maatschappelijke organisaties gebruiken onze methoden in 2026 • Internationale standaard verwijst naar ons werk voor 2028 • Beschikbaar stellen van tools of instrumenten voor ontwikkelaars en beleidsmakers
Organisatorische gezondheid	• Verbreden van financieringsbronnen vanaf 2026 • Minimaal 90% van begrote activiteiten uitgevoerd per jaar dienen publiekelijk belang

Tabel 1: Kwaliteitsindicatoren 2025-2028

3 Huidige Situatie en Activiteiten

Aithos is opgericht in 2025 omdat er geen organisatie bestond die zich richtte op het fundamentele probleem van AI-waardenmonopolie en de depolitisering van technologische keuzes. Terwijl meerdere organisaties zich bezighouden met AI-veiligheid, discriminatie-preventie, of technische standaarden, richt niemand zich specifiek op de vraag: hoe brengen we AI-waarden naar de publieke sfeer waar verschillende perspectieven kunnen delibereren?

3.1 Kernactiviteiten 2025-2028

AI-Waarden Zichtbaar Maken

Aithos geeft de prioriteiten van AI-systemen weer in begrijpelijke beschrijvingen en grafieken. We onderzoeken welke waarden AI-systemen daadwerkelijk hanteren door ze in gecontroleerde omgevingen waarin AI-modellen worden geconfronteerd met realistische ethische dilemma's te plaatsen, en hun keuzes systematisch te analyseren. Dit omvat het in kaart brengen van AI-gedrag wanneer verschillende waarden botsen en het identificeren van patronen in hun besluitvorming. We evalueren ook hoe consistent AI-systemen zijn in hun waardenkeuzes, hoe goed ze hun redeneringen kunnen uitleggen, en of hun gedrag overeenkomt met regelgeving.

Onderzoek naar Contextafhankelijkheid

Door verschillende AI-modellen te testen, en te variëren in contextuele factoren zoals functie, informatie over impact, tijdsdruk, autoriteit en publieke zichtbaarheid, onderzoeken we hoe deze elementen AI-gedrag beïnvloeden. We doen dit in een realistische context, om de problemen die zich in praktische situaties kunnen voordoen en de meest effectieve manier om systemen aan te sturen te identificeren.

Publieke Transparantie en Kennisdeling

Transparantie staat centraal in onze missie om AI-waarden te democratiseren. We publiceren onze bevindingen over de waardensystemen van grote AI-modellen openbaar en ontwikkelen tools die gebruikers helpen begrijpen welke waarden hun AI-systemen hanteren. Onze onderzoeksmethoden en resultaten delen we via wetenschappelijke publicaties om bij te dragen aan het bredere academische en maatschappelijke debat.

Gepersonaliseerde Afstemming

We onderzoeken hoe mensen en organisaties gekoppeld kunnen worden aan AI-systemen die hun specifieke waarden respecteren, zonder dat één waardensysteem andere domineert. Dit vereist de ontwikkeling van geavanceerde technieken die verder gaan dan gangbare trainingsmethodes en die echte ethische complexiteit kunnen hanteren. We werken aan pluralistische configuratiemogelijkheden waarmee AI-systemen binnen juridische kaders kunnen worden afgestemd op diverse behoeften en prioriteiten.

Wetenschappelijk Onderzoek

Ons onderzoek combineert inzichten uit filosofie, sociale wetenschappen, informatica en bestuurskunde om een holistisch begrip van AI-waarden te ontwikkelen. We voeren empirische studies uit naar hoe waardenkeuzes in de

praktijk functioneren en dragen bij aan de theoretische ontwikkeling van AI-ethiek en pluralisme. Door nieuwe methoden te ontwikkelen verleggen we de grenzen van hoe we AI-waarden kunnen begrijpen, meten en beïnvloeden.

3.2 Praktijkvoorbeeld: Zichtbaar Maken van AI-Waarden in Klantenservice

Een recent voorbeeld illustreert het probleem waar Aithos zich over ontfermt: een chatbot van luchtvaartmaatschappij Air Canada gaf een klant onjuiste informatie over rouwtarieven. De klant kocht op basis van deze informatie een duur ticket, maar kreeg later zijn geld niet terug. Een rechter oordeelde dat Air Canada verantwoordelijk was voor de foutieve informatie van hun AI-systeem. Dit geval toont aan dat AI-systemen nu al consequenties hebben voor gewone burgers, en dat de waardenkeuzes die in deze systemen zijn ingebouwd grondig moeten worden beschouwd, en moeten voldoen aan de doelstellingen van de organisaties die ze inzetten.

Een concreet project voor 2025-2026 richt zich op het blootleggen van de verborgen waardenkeuzes die AI-chatbots maken in klantenservice. Deze systemen maken voortdurend ethische afwegingen die grote impact hebben op gebruikers, maar deze keuzes zijn onzichtbaar voor zowel klanten als bedrijven. Ons onderzoek werkt als volgt: We creëren een gecontroleerde simulatie-omgevingen waarin AI-systemen worden geconfronteerd met moeilijke ethische dilemma's. Bijvoorbeeld:

- Een klant vraagt om een product te retourneren en vertelt een emotioneel verhaal over financiële problemen – moet het systeem coulanter zijn of juist strikt de regels volgen?
- Twee klanten hebben vergelijkbare klachten, maar één dreigt met juridische stappen – behandelt het systeem hen gelijk of geeft het prioriteit aan de dreiging?
- Een bedrijf geeft instructies die de wet zouden breken door privacy te schenden. Volgt het systeem de instructies of wettelijke regels?
- Is er sprake van vooroordelen in het behandelen van verschillende klanten met vergelijkbare situaties?

Door systematisch te variëren in factoren zoals tijdsdruk, schade, emotionele context, juridische risico's en publieke zichtbaarheid, kunnen we in kaart brengen welke waarden het AI-systeem werkelijk hanteert. We visualiseren dit als een waardenkaart die laat zien hoe het systeem afweegt tussen bijvoorbeeld:

- Klanttevredenheid versus bedrijfsregels
- Efficiëntie versus zorgvuldigheid
- Gelijke behandeling versus contextgevoeligheid
- Risicomijding versus hulpvaardigheid

De resultaten maken we op twee manieren beschikbaar:

- **Voor het publiek:** Begrijpelijke rapporten die laten zien welke waarden verschillende bedrijven hanteren in hun AI-systemen, zodat consumenten geïnformeerde keuzes kunnen maken
- **Voor organisaties:** Tools waarmee bedrijven hun AI-systemen kunnen afstemmen op hun gewenste waarden en kunnen controleren of deze consistent worden toegepast

Dit project toont onze kernbenadering: in plaats van te bepalen welke waarden 'juist' zijn, maken we zichtbaar welke waarden verschillende systemen hanteren, zodat verschillende stakeholders hierover kunnen delibereren en geïnformeerde keuzes kunnen maken.

4 Toekomstige Ontwikkelingen

AI-systemen worden steeds zelfstandiger en nemen beslissingen die vroeger door mensen werden genomen. Deze ontwikkeling vergroot de urgentie van Aithos' werk om AI-waarden zichtbaar en bespreekbaar te maken.

Meer Zelfstandige AI-Systemen: AI evolueert van systemen die vragen beantwoorden naar systemen die zelfstandig taken uitvoeren. Een AI-assistent die alleen maar e-mails beantwoordt, maakt andere waardenkeuzes dan een systeem dat zelfstandig afspraken plant, leveranciers selecteert, of financiële beslissingen neemt. Hoe zelfstandiger deze systemen worden, hoe belangrijker het wordt dat hun waardenkeuzes overeenkomen met die van de mensen en organisaties die ze dienen.

AI in Kritieke Besluitvorming: We zien AI-systemen steeds vaker worden ingezet bij belangrijke beslissingen: welke sollicitanten worden uitgenodigd, welke medische behandelingen worden voorgesteld, hoe budgetten worden verdeeld. Momenteel worden deze waardenkeuzes gemaakt door de bedrijven die de AI ontwikkelen, zonder input van de gemeenschappen die ermee te maken krijgen, maar de EU AI Act zal transparantie in dit soort systemen een vereiste maken. Aithos verwacht dat dit zal leiden tot groeiende spanning tussen verschillende groepen over hoe AI-systemen zouden moeten functioneren.

Schaalvergroting van Waardenkeuzes: Een menselijke klantenservice medewerker maakt misschien honderd beslissingen per dag. Een AI-systeem kan er miljoenen maken. Kleine verschillen in waardenkeuzes – bijvoorbeeld of je meer prioriteit geeft aan klanttevredenheid of aan het volgen van regels – worden enorm vergroot als ze door AI-systemen worden toegepast. Dit maakt het belangrijker dat deze keuzes bewust worden gemaakt in plaats van per ongeluk.

Verbreiding van functionaliteit Naarmate de capaciteiten van AI-systemen groeien zullen ze meer worden ingezet, tot het punt waarop een gemiddeld mens constant omgaat met AI-systemen: om spullen te kopen, informatie te vinden, en afspraken te maken. Dit maakt gebruikers gevoeliger voor misinformatie en manipulatie.

Regulatoire Druk: De EU AI Act en vergelijkbare wetgeving in andere landen verplichten organisaties steeds meer om verantwoording af te leggen over hun AI-systemen. Dit creëert vraag naar de tools die Aithos ontwikkelt om AI-waarden zichtbaar te maken. Tegelijkertijd leiden deze regelgevingen tot de noodzaak om transparantie en verantwoording te realiseren, waar het in kaart brengen van individuele waardesystemen een belangrijk onderdeel van zou kunnen zijn.

Maatschappelijke Bewustwording: Burgers worden zich steeds meer bewust van de rol van AI in hun dagelijkse leven. Dit leidt tot meer vragen over hoe deze systemen werken en wiens waarden ze vertegenwoordigen. Nederlandse AI governance is zelfs open-source om dit te kunnen faciliteren. Aithos verwacht een groeiende vraag naar transparantie en democratische inspraak in AI-governance.

Internationale Ontwikkelingen: Verschillende landen en regio's ontwikkelen verschillende benaderingen tot AI-governance. Deze diversiteit biedt kansen voor het testen van pluralistische methoden, maar creëert ook uitdagingen voor organisaties die internationaal opereren.

4.1 Positionering Aithos

Proactieve Tool-ontwikkeling: Naarmate AI-systemen zelfstandiger worden, ontwikkelt Aithos tools die geschikt zijn voor complexere waardenkeuzes. In plaats van te wachten tot problemen ontstaan, anticiperen we op de behoeften van organisaties die verantwoordelijk AI willen implementeren.

Samenwerking met Regelgevers: Aithos werkt samen met beleidsmakers om ervoor te zorgen dat nieuwe regelgeving ruimte laat voor waardendiversiteit. We adviseren over hoe AI-governance kan worden ingericht zonder dat één specifieke visie wordt opgelegd aan alle sectoren en gemeenschappen, en mengen ons in het publieke debat om bewust te blijven van meerdere meningen.

Schaalbaarheid van Methoden: Aithos ontwikkelt methoden die zowel geschikt zijn voor kleine als voor grootschalige AI-implementaties, en simpele modellen even goed kunnen evalueren als de meest geavanceerde op de markt. Onze tools worden ontworpen om te groeien met de complexiteit van AI-systemen.

Ontwikkeling van Beschermende AI Blootstelling aan een veelvoud aan AI maakt mensen ontvankelijk voor misinformatie en manipulatie. Aithos doet onderzoek naar de mogelijkheid om AI-afstemming te personaliseren, zodat het contact met andere AI-systemen grotendeels ondersteund kan worden door een eigen AI-assistent die de persoonlijke waarden en het belang van de gebruiker uitdraagt, en deze beschermt tegen ongewenste invloed.

Internationale Kennisdeling: Aithos deelt kennis en methoden internationaal om te voorkomen dat waardenmonopolies zich over landsgrenzen heen uitbreiden. We werken samen met gelijkgestemde organisaties om pluralistische benaderingen wereldwijd te promoten.

Publieke Educatie: Naarmate maatschappelijke bewustwording groeit, investeert Aithos in het toegankelijk maken van kennis over AI-waarden voor een breed publiek. We ontwikkelen educatiematerialen die mensen helpen begrijpen hoe AI-systemen werken en hoe ze hierop invloed kunnen uitoefenen.

Flexibele Organisatiestructuur: Aithos houdt haar organisatie flexibel om snel te kunnen reageren op nieuwe ontwikkelingen. We bouwen netwerken op met diverse partners zodat we verschillende expertises kunnen mobiliseren als dat nodig is.

Deze ontwikkelingen versterken elkaar en maken Aithos' werk steeds relevanter. Door proactief te reageren op deze trends, positioneert Aithos zich als een essentiële partner voor organisaties en gemeenschappen die AI willen implementeren op een manier die hun waarden respecteert.

5 Organisatiestructuur

Contactgegevens:

- KvK-nummer: 97562491
- RSIN: 868111788
- Contact: info@aithos.org, Keizersgracht 62, 1015 CS Amsterdam

Aithos streeft naar ANBI-status om haar democratiseringsmissie te ondersteunen met belastingvrije schenkingen.

5.1 Bestuur

Het bestuur bestaat uit drie onafhankelijke leden met complementaire expertise:

- **Voorzitter:** dhr. Lennard Zwart
Verantwoordelijk voor leiding van het bestuur en algemene vergaderingen, en vertegenwoordigen van de organisatie naar binnen en naar buiten
- **Secretaris:** dhr. Sicco Pier van Gosliga
Verantwoordelijk voor het schrijven van brieven namens de stichting, verslagen, en notulen van vergaderingen.
- **Penningmeester:** dhr. Maarten Timmerman
Verantwoordelijk voor het bijhouden van kas- en bankboek, en beheren van de bankrekeningen, maken van een financieel jaarverslag, en bewaken van de begroting.

Bestuursleden zijn onbezoldigd en ontvangen alleen onkostenvergoedingen. Het bestuur vergadert minimaal vier keer per jaar. De bestuurstaken kunnen gedelegeerd worden aan de uitvoerende organisatie, maar het bestuur houdt eindverantwoordelijkheid voor en toezicht op alle activiteiten.

5.2 Uitvoerende Organisatie

Aithos begint met een compacte maar gespecialiseerde organisatie:

- **Uitvoerend Directeur:** dhr. Daan Henselmans (circa 0.6 FTE)
Verantwoordelijk voor onderzoeksstrategie, organisatie, en coördinatie van technische waarden-analyse
- **Onderzoeker:** dhr. Derck Prinzhorn (circa 0.6 FTE) Verantwoordelijk voor technische ontwikkeling van tools voor empirisch onderzoek naar AI-waarden

Bij voldoende financiering wordt het team uitgebreid met specialisten voor democratische facilitatie, statistiek, en interdisciplinaire wetenschappen.

5.3 Samenwerkingsnetwerk

Aithos heeft het doel samen te werken met diverse partners met relevante expertise:

- Nederlandse universiteiten en kennisinstellingen
- Overheidsorganisaties op verschillende niveaus
- Maatschappelijke organisaties met verschillende waardensystemen
- Internationale partners in pluralistische AI-governance
- Lokale gemeenschappen en belangengroeperingen

6 Integriteitsprotocol

6.1 Conflictenbeleid (Overgangsfase)

Gezien de initiële verbondenheid tussen het bestuur en Ascending.ai, hanteert Aithos het volgende tijdelijke kader:

- **Transparantie:**

Alle bestuursfuncties bij Ascending.ai worden gepubliceerd in het jaarverslag. Donaties worden per transactie openbaar gemaakt

- **Besluitvorming:**

Voor eventuele besluiten die effect hebben op Ascending.ai wordt een externe goedkeuringsprocedure gevolgd:

1. Schriftelijke motivering van het besluit
2. Toetsing door een onafhankelijk jurist
3. Publieke vastlegging van de afweging

- **Transitieplan:**

Binnen 18 maanden wordt ten minste één bestuurspositie vervuld door een persoon zonder banden met Ascending.ai

Financiële afhankelijkheid van Ascending.ai wordt binnen 3 jaar teruggebracht tot <50% van het totale budget

6.2 Ethische Principes voor Bestuur:

Onafhankelijkheid: Bestuursleden handelen uitsluitend in het belang van Aithos' missie, niet vanuit persoonlijke of externe belangen.

Transparantie: Financiële stromen en besluitvormingsprocessen zijn traceerbaar en controleerbaar.

Pluralisme: Geen enkel waardensysteem krijgt voorrang in onderzoek of tools, ongeacht financieringsbronnen.

Wetenschappelijke Integriteit: Onderzoekresultaten worden nooit gemanipuleerd ten behoeve van sponsors.

6.3 Klokkenluidersregeling:

- Een extern vertrouwenspersoon (geen bestuurslid) behandelt meldingen via integrity@aithos.org
- Anonieme meldingen zijn mogelijk via een beveiligd online formulier op de website
- Bescherming van klokkenluiders gegarandeerd volgens de Wet bescherming klokkenluiders
- Externe controle bij ernstige beschuldigingen, met geanonimiseerde publicatie van bevindingen

7 Financiële Planning

Aithos is opgericht op 20 juni 2025, als stichting zonder winstoogmerk, en hanteert een transparante en onafhankelijke financieringsstrategie die onze pluralistische missie ondersteunt.

Startfinanciering: Aithos ontvangt initiële financiering van €200.000 van Ascending.ai, een Nederlands bedrijf dat pluralistische AI-ontwikkeling ondersteunt. Deze middelen worden ingezet voor onderzoek dat democratische deliberatie over AI-waarden faciliteert.

Jaarbegroting 2025-2028:

- **Personeelskosten** (onderzoek en facilitatie)
 - 2 ZZP'ers €120.000
 - Inhuur expertise € 16.000
 - BTW-kosten € 28.560
- **Marketing en evenementen** € 10.000
- **Onderzoekskosten** (AI, API) € 4.000
- **Administratie** (boekhouding) € 5.000
- **Operationele kosten** (kantoor) € 6.000
- **Reiskosten** (conferenties) € 5.000

-
- **Totaal jaarbudget:** €194.560

Toekomstige Financiering:

- **Partnerschap:** Bij het succesvol behalen van de doelen verwacht Aithos verdere schenkingen van Ascending.ai te ontvangen.
- **Onderzoekssubsidies:** NWO-beurzen voor democratische innovatie en pluralisme-onderzoek
- **Pluralistische en Lange-Termijn Filantropie:** Schenkingen van internationale organisaties die waardendiversiteit en welzijn op de lange termijn ondersteunen
- **Participatieve Partnerships:** Samenwerking met initiatieven voor democratische innovatie

Financieel Beheer:

- Alle middelen worden uitsluitend gebruikt voor bevordering van pluraliteit, democratisering, en onderzoek naar AI-waarden van algemeen belang
- Financiële reserve van 6 maanden voor continuïteit
- Jaarlijkse accountantscontrole conform ANBI-eisen
- Volledige transparantie: alle financiële informatie wordt jaarlijks gepubliceerd

Maatschappelijke Waarde

Elke euro die Aithos ontvangt wordt ingezet voor het inzichtelijk maken van AI-waarden en het representeren van legitieme waardenverschillen. Onze resultaten zijn altijd openbaar beschikbaar en dienen geen enkel specifiek waardensysteem boven anderen.

Liquidatie

In het geval dat Aithos wordt opgeheven met een positief liquidatiesaldo, wordt het overige saldo gedoneerd aan een andere Nederlandse ANBI-organisatie met vergelijkbare doelen en maatschappelijke waarde, en/of een buitenlandse instelling met een soortgelijk doel die voor minimaal 90% algemeen belang nastreeft.