GRATIS WERKDOCUMENT · RUDYJELLESMA.NL

Het Ai-stoplicht

Wat mag Ai zelf in jouw bedrijf? Het groen/geel/rood-model voor Ai-autonomie.

Rudy Jellesma

rudyjellesma.nl

Werkdocument · 2026

Waarom dit kader?

Elke ondernemer die met Ai-agents begint, loopt tegen dezelfde vraag aan: *wat mag dat ding eigenlijk zelf doen?* Mag een Ai-assistent zelfstandig mails beantwoorden? Een factuur inboeken? Een wijziging op je website live zetten?

De meeste bedrijven beantwoorden die vraag pas nadat het is misgegaan. Dan blijkt dat de Ai een klant een verkeerd bedrag heeft gemaild, of dat een "kleine fix" de website een middag offline heeft gezet. Het omgekeerde komt net zo vaak voor: uit angst mag de Ai helemaal níks zelf, en dan levert de hele investering weinig meer op dan een dure chatbox.

Ik draai zelf tientallen Ai-agents die 24/7 werken (monitoring, code-controle, boekhoudcontrole, mail-triage), naast al het andere werk in mijn bedrijven. Dat gaat alleen goed omdat ik van tevoren, op papier, heb vastgelegd wat elke agent wél en níet zelfstandig mag. Wat ik bij grote techorganisaties terugzie wijst dezelfde kant op: vrijwel niemand draait Ai volledig autonoom in productie. De norm is een adviserende Ai, met een handvol duidelijk afgebakende categorieën waarin de Ai wél zelf mag handelen.

Dat model is verrassend simpel uit te leggen. Het is een stoplicht.

De belangrijkste regel van dit hele document staat hieronder, dus die krijg je meteen:

> **Schrijf de grenzen op vóórdat je bouwt of koopt. Niet erna.**

Een Ai-tool zonder vooraf vastgelegde grenzen krijgt zijn grenzen vanzelf: door incidenten. Dit werkdocument helpt je die grenzen in één middag op papier te zetten.

Het stoplichtmodel: groen, geel, rood

Geef je Ai een rijbewijs met drie categorieën. Elke actie die een Ai-systeem in jouw bedrijf kan uitvoeren, valt in precies één van deze drie kleuren.

Groen: de Ai doet het zelf, zonder melding

Acties die **omkeerbaar** zijn en een **kleine impact** hebben als het misgaat. Denk aan:

- Informatie **lezen en samenvatten**: mail scannen, rapportages maken, logbestanden doorzoeken
- **Concepten voorbereiden**: een antwoordmail als klad klaarzetten, een offerte-opzet maken
- **Interne analyses**: "welke facturen staan langer dan 30 dagen open?"
- **Controles draaien** die niets wijzigen: is de website bereikbaar, klopt de voorraadtelling?

De toets voor groen: *als de Ai dit honderd keer per dag fout doet, is er dan blijvende schade?* Nee? Dan is het groen.

Geel: de Ai doet het én meldt het

Acties die de Ai zelfstandig uitvoert, maar waarvan jij (of een collega) **achteraf een korte melding** krijgt. Geel is voor acties die wél iets veranderen, maar **eenvoudig terug te draaien** zijn:

- Een intern document of conceptpagina aanpassen

- Een taak op de planning zetten of verplaatsen
- Een vastgelopen intern proces herstarten
- Een voorstel-boeking klaarzetten in de boekhouding (nog niet definitief boeken)

De melding is geen formaliteit, het is jouw steekproef. Wie de gele meldingen nooit leest, heeft in de praktijk geen geel meer, maar een tweede groen.

Rood: nooit zonder mens

Acties die **onomkeerbaar** zijn, **geld kosten**, **naar buiten gaan** of **juridische gevolgen** hebben. De Ai mag ze voorbereiden tot en met de laatste stap, maar een mens drukt op de knop:

- Alles wat naar een **klant of externe partij** wordt verstuurd
- **Betalingen**, definitieve boekingen, aangiftes
- **Verwijderen** van data die je niet kunt herstellen
- Wijzigingen **live zetten** op systemen waar klanten van afhankelijk zijn
- Contracten, toezeggingen, prijsafspraken

Mijn hardste persoonlijke regel in deze categorie: **laat een Ai nooit zelf op de productie-knop drukken.** Een wijziging live zetten op een systeem waar klanten op draaien is een eenrichtingsdeur. In mijn eigen opzet mag een Ai-agent bouwen en testen in een oefenomgeving zoveel hij wil, maar de stap naar “live” loopt via een vast, voorspelbaar script met ingebouwde gezondheidscontrole en automatische terugdraai-optie, plus een menselijk akkoord zodra er betalende klanten geraakt kunnen worden.

De autonomie-tier verschilt per proces, en dat is de bedoeling

Eén bedrijf, verschillende regimes. Mijn hobbyproject mag zichzelf grotendeels beheren; alles wat met facturatie te maken heeft mag dat uitdrukkelijk niet. Dat is geen inconsistentie, dat is precies waar het Ai-stoplicht voor dient. Hoe groter de schade bij een fout, hoe kleiner de autonomie.

Extra waarborgen voor “zelf-herstellende” acties

Wil je een Ai ook laten **ingrijpen** (bijvoorbeeld automatisch een vastgelopen systeem herstarten), hanteer dan vier voorwaarden die uit de praktijk van grote techbedrijven komen, en die ik zelf op de harde manier heb geleerd:

1. **Betrouwbaar signaal**: de Ai grijpt alleen in als het probleem écht is vastgesteld, niet op een vermoeden.
2. **Herhaalbaar zonder schade**: de actie twee keer uitvoeren mag geen extra schade geven.
3. **Kleine impact**: één systeem tegelijk, nooit “herstart alles”.
4. **Omkeerbaar**: het terugdraai-pad is vooraf bekend.

Plus drie vangrails: maximaal 2-3 pogingen (daarna stoppen en escaleren naar een mens), controleer vóór de actie of het probleem nog bestaat, en controleer ná de actie of het probleem echt weg is, niet alleen of “het commando is gelukt”.

De drie anti-patronen: zo gaat het alsnog mis

Een stoplicht op papier is stap één. Deze drie sluipwegen maken het in de praktijk kapot.

1. Yellow-creep

De gevaarlijkste: **veel kleine gele acties die samen één rode actie vormen.** Elke stap is op zichzelf onschuldig (een instelling hier, een documentje daar, een herstart erbij), maar het totaal is een ingreep die je nooit als één geheel zou hebben goedgekeurd. Vijf gele acties op hetzelfde bedrijfskritische systeem zijn de facto rood.

De regel: meerdere gele acties op hetzelfde kritieke proces binnen korte tijd = behandelen als rood. Herken het patroon en laat de Ai (of jezelf) op dat moment alsnog om akkoord vragen.

2. “De agent zei dat het klaar was”

Het tweede anti-patroon is menselijk: **de Ai als excuus gebruiken.** “De Ai heeft het gecontroleerd.” “Het systeem gaf groen licht.” Wie een gele melding afvinkt zonder te kijken, of een rood akkoord geeft zonder het voorstel te lezen, heeft de verantwoordelijkheid niet gedelegeerd, hij heeft haar laten vallen.

Dit is geen theoretisch risico. Ik heb meegemaakt dat een complete testsuite groen was, terwijl de wijziging het systeem in de praktijk stilletjes zou hebben gebroken. Een fout van de klasse “**slaagt-maar-fout**”: alles ziet er geslaagd uit, maar het resultaat klopt niet. De Ai-agent in kwestie deed toen overigens precies het goede: hij weigerde de grote wijziging in één keer door te voeren, bouwde in plaats daarvan ~15 kleine, elk afzonderlijk omkeerbare voorbereidingsstappen, en legde de definitieve omschakeling expliciet bij de mens neer. Dát is het gedrag dat je wilt afdwingen: Ai levert veilige, kleine stappen; het onomkeerbare besluit blijft bij jou.

De regel: wie akkoord geeft, kijkt zelf naar het resultaat. Steekproefsgewijs mag; nooit kijken mag niet.

3. Stuck-loops

Het derde patroon: **een Ai die blijft proberen.** Vijf keer dezelfde mislukte aanpak met kleine variaties, of erger: een agent die eindeloos werk produceert dat niemand afmaakt. Ik heb een agent gehad die keurig verbetervoorstellen bleef aanmaken zonder dat iets ooit werd afgerond: binnen twee weken lag er een stapel van ruim een dozijn open voorstellen. Veel activiteit, nul effect, en elke poging kost geld.

De regels: na 2 à 3 mislukte pogingen stopt de Ai en meldt het probleem. Meet het **effect** van je Ai (afgeronde taken), niet de **activiteit** (gestarte taken). En check vóór elke nieuwe taak of dezelfde taak niet al openstaat.

De invulmatrix

Hieronder de matrix met voorbeeldacties per bedrijfsproces. Dit zijn **startwaarden**: streep door, verschuif en vul aan tot hij bij jouw bedrijf past. De laatste kolom is verplicht: elke rij heeft een mens met naam nodig.

Proces	Groen (Ai doet zelf)	Geel (Ai doet + meldt)	Rood (mens beslist)	Mens-in-de-loop
Mail	Inkomende mail lezen, labelen, samenvatten; ochtendoverzicht maken	Conceptantwoord klaarzetten in de map “concepten”; afspraak-voorstel inplannen	Mail daadwerkelijk versturen naar klant of externe partij	Eigenaar / office manager

Proces	Groen (Ai doet zelf)	Geel (Ai doet + meldt)	Rood (mens beslist)	Mens-in-de-loop
Boekhouding	Facturen uitlezen; banktransacties matchen aan facturen; afwijkingen signaleren	Voorstel-boekingen klaarzetten (grootboek + btw-code) in een controlebatch	Definitief boeken; betalen; btw-aangifte indienen; iets fiscaals bij twijfel	Eigenaar / boekhouder
Website	Bezoekcijfers en fouten monitoren; verbetervoorstellen doen; teksten als concept schrijven	Wijziging doorvoeren op de test-/oefenomgeving	Wijziging live zetten; pagina's verwijderen; prijzen aanpassen	Eigenaar / webbeheerder
Klantcommunicatie	Vraag van klant analyseren; antwoord voorbereiden met bronvermelding; klantdossier samenvatten	Interne notitie bij het klantdossier; collega intern informeren	Elk bericht dat de klant daadwerkelijk ontvangt; toezeggingen; compensaties	Accountverantwoordelijke
Deploys / IT	Systemen monitoren; logbestanden analyseren; storing diagnosticeren	Vastgelopen intern systeem herstarten (één tegelijk, max 2 pogingen, daarna melden)	Live zetten op productie; data verwijderen; toegangsrechten wijzigen; alles met wachtwoorden/sleutels	IT-verantwoordelijke
Data-correcties	Inconsistenties opsporen en rapporteren ("deze 12 records wijken af")	Correctievoorstel klaarzetten mét terugdraai-plan per record	Correcties massaal doorvoeren; records verwijderen; klantdata wijzigen	Proceseigenaar

Drie leesregels bij deze matrix:

- **Twijfel je tussen twee kleuren? Kies de strengste.** Opschuiven naar meer autonomie kan later altijd, op basis van bewezen prestaties.
- **"Naar buiten" is altijd rood.** Wat het bedrijf verlaat (mail, bericht, betaling, publicatie) kan niet worden teruggeroepen.
- **De mens-in-de-loop is een naam, geen afdeling.** "Iemand van kantoor" betekent in de praktijk: niemand.

In 10 stappen ingevoerd, morgen te beginnen

1. **Inventariseer wat er al draait.** Welke Ai-tools en automatiseringen gebruikt je bedrijf nu al, inclusief de dingen die medewerkers zelf zijn gaan gebruiken? Je kunt geen grenzen stellen aan wat je niet kent.
2. **Kies de zes processen die ertoe doen.** Begin met de rijen uit de matrix hierboven en schrap of vervang wat niet van toepassing is. Zes rijen zijn genoeg voor versie 1.

3. **Vul per proces de drie kleuren in.** Gebruik de toetsvragen: omkeerbaar + kleine impact = groen; omkeerbaar maar merkbaar = geel; onomkeerbaar, extern of geld = rood.
4. **Zet bij elke rij een naam.** Wie leest de gele meldingen? Wie geeft rode akkoorden? Eén persoon per rij, met een vervanger voor vakanties.
5. **Leg de matrix vast op één vindbare plek.** Eén A4 of één pagina in je bedrijfswiki. Niet in iemands hoofd, niet in drie versies in de mail.
6. **Regel de logging.** Elke Ai-actie, ook een groene, moet terug te vinden zijn: wat is er gedaan, wanneer, en waarom. Zonder logboek kun je nooit vaststellen wat er gebeurd is, en dus ook nooit verantwoord autonomie uitbreiden. De striktere variant, die ik zelf gebruik: een vaste controlelaag die élke actie vooraf tegen de matrix houdt en bij twijfel blokkeert. En als die controlelaag zelf uitvalt, staat álles op rood (“fail-closed”).
7. **Bouw het gele meldkanaal.** Eén plek (teamchat, dagelijkse digest-mail) waar alle gele meldingen samenkomen. Belangrijk: gebundeld, niet per gebeurtenis, anders leert iedereen binnen een week de meldingen te negeren.
8. **Draai twee weken proef in de strengste stand.** Zet in het begin méér op rood dan je denkt nodig te hebben. Noteer welke rode akkoorden pure formaliteit blijken: dat zijn je kandidaten voor geel.
9. **Evalueer maandelijks, op effect, niet op activiteit.** Welke gele acties gingen 30 dagen foutloos? Die mogen richting groen. Waar ontstond bijna-schade? Die schuift terug naar rood. De matrix is een levend document.
10. **Bespreek de anti-patronen met het team.** Yellow-creep, “de Ai zei dat het goed was” en stuck-loops zijn menselijke valkuilen, geen technische. Eén teamoverleg van een half uur waarin je ze benoemt, voorkomt meer schade dan welke tool ook.

Tot slot

Het Ai-stoplicht is geen rem op Ai; het is juist wat verantwoord gasgeven mogelijk maakt. Wie zijn grenzen kent, durft binnen die grenzen veel meer aan de Ai over te laten. En wie ze niet kent, komt er via een incident achter waar ze hadden moeten liggen.

Begin klein: zes processen, drie kleuren, één naam per rij, één middag werk.

Meer praktijkverhalen over Ai in het MKB, inclusief de missers, lees je op **rudyjellesma.nl**. Vragen over dit werkdocument of hulp nodig bij het invullen voor jouw bedrijf? Ook daarvoor kun je daar terecht.

Over de auteur

Rudy Jellesma is ondernemer en CTO. Hij bouwt en beheert AI-systemen die dag en nacht meewerken in zijn eigen bedrijven, van monitoring en codecontrole tot boekhouding en mail-afhandeling. Op rudyjellesma.nl deelt hij wat daarbij werkt en wat misgaat, steeds vertaald naar de praktijk van het MKB.

Verder lezen

Praktijkverhalen, tips en stappenplannen over AI voor het MKB vind je op `rudyjellesma.nl`.

Andere gratis downloads:

- Ai-startklaar-checklist voor MKB
- Ai-kostenwijzer voor MKB
- BTW-checklist voor Ai-abonnementen
- Ai-boekhouden met akkoord-gate
- Beveilig je Ai-agent in 7 stappen
- Start je bedrijfs-kennisbank voor Ai in 1 dag