

Aanbevelingen AI in de advocatuur

Herziene versie d.d. november 2026

Aanbevelingen AI in de advocatuur

Herziene versie d.d. november 2026

Over deze aanbevelingen

Deze aanbevelingen zijn richtinggevend en bedoeld als praktische ondersteuning voor advocaten om verantwoord AI-gebruik te bevorderen. De advocaat blijft dominus litis. De aanbevelingen zijn op zichzelf geen dwingende tuchtrechtelijke of wettelijke eisen[1]. Proportionaliteit, subsidiariteit, het toepasselijk wet- en regelgevingskader blijven bij de toepassing ervan leidend. Advocaten worden in het bijzonder gewezen op de verplichtingen die voor hen gelden onder de AI Verordening (EU 2024/1689) ten aanzien van het gebruik van AI-systemen en de verplichtingen die gelden op basis van de AVG (EU 2016/679) ten aanzien van de verwerking van persoonsgegevens.

Dit is een levend document. In deze tijd van snelle technologische ontwikkelingen dienen deze aanbevelingen als vorm van bewustwording, opleiding en risicomitigatie om acute risico's binnen de context van ons beroep te helpen beperken. Zij zullen op basis van voortschrijdend inzicht en technologische ontwikkeling periodiek worden aangepast.

Het doel van deze aanbevelingen is de frictie met de kernwaarden van de advocatuur (deskundigheid, vertrouwelijkheid, onafhankelijkheid, integriteit en partijdigheid) bloot te leggen en handvatten te bieden voor beoordeling. Ze zijn geen vervanging van een eigen situatie-specifieke beoordeling van compliance met alle toepasselijke wet- en regelgeving en/of informatiebeveiligingsnormen en ook niet als zodanig bedoeld.

Deze aanbevelingen zijn van toepassing op alle vormen van AI die advocaten mogelijk inzetten, waaronder generatieve AI (tekst, beeld, audio, video), AI-ondersteunde zoek- en analysediensten, en agentic AI-systemen.

De aanbevelingen zijn ingedeeld in vijf onderwerpen. Een lijst met veelgebruikte basisbegrippen is opgenomen in de bijlage aan het einde van dit document.

A Beleid ten aanzien van inkoop en gebruik van AI
B Bekwaamheid en gebruik
C Informatiebeveiliging
D Transparantie en verantwoording
E Partijdigheid en maatschappelijke impact

[1] Omwille van de leesbaarheid zijn de aanbevelingen veelal in de gebiedende wijs geformuleerd, bijvoorbeeld 'stel vast', 'controleer' en 'zorg voor'. Deze formulering maakt de aanbevelingen niet bindend en schept op zichzelf geen zelfstandige verplichtingen. De advocaat bepaalt welke toepassing in de concrete situatie passend is, rekening houdend met de aard van de werkzaamheden, de risico's en de toepasselijke wet- en regelgeving en beroepsnormen. Waar een aanbeveling aansluit bij een bestaande verplichting, vloeit het bindende karakter voort uit die verplichting, niet uit dit document.

A Beleid ten aanzien van inkoop en gebruik van AI

A.1 Stel een risicogebaseerd, kantoorbreed AI-beleid op en dwing zoveel mogelijk technisch en organisatorisch naleving af

1 Aanbeveling

Ieder advocatenkantoor beschikt over een schriftelijk kantoorbreed AI-beleid dat verduidelijkt hoe hun advocaten AI verantwoord en effectief benutten terwijl de kernwaarden van de advocatuur - in het bijzonder de geheimhoudingsplicht, partijdigheid en deskundigheid - gewaarborgd blijven. Wanneer opportuun (bijvoorbeeld om gevoelige bedrijfsinformatie te beschermen) kunnen kantoren ervoor kiezen om een extern en een intern AI-beleid te hanteren waarbij het externe beleid vooral inzicht geeft in de gevolgen en waarborgen voor cliënten en het interne beleid meer inzicht geeft in concrete risicoafwegingen en mitigerende maatregelen. Het AI-beleid is risicogebaseerd en houdt rekening met zowel de risico's van de AI-tool in algemene zin (i.e. ongeacht de gegevens die verwerkt worden) als de risico's van concrete gebruik van die AI-tool (i.e. gelet op het soort gebruik en de gegevens die daarmee verwerkt worden).

2 Toelichting

Het ontbreken van heldere kaders leidt in de praktijk niet tot het uitblijven van AI-gebruik, maar tot medewerkers die op eigen initiatief consumententools inzetten zonder de contractuele of technische waarborgen te kennen. Aanbevolen wordt om in het AI-beleid ten minste de volgende onderwerpen te adresseren:

(1) Welke contractuele en technische waarborgen (zero data retention, geheimhoudingsbedingen, serverlocatie, encryptie, en waar relevant anonimisering of pseudonimisering) noodzakelijk zijn vóórdat zaakgegevens aan een AI-systeem worden toevertrouwd. In dat kader moet benadrukt worden dat het verschoningsrecht en het beroepsgeheim verder strekken dan de algemene privacywetgeving;

(2) Een whitelist van goedgekeurde tools met afgebakende toegestane en uitgesloten of verzwaarde soorten toepassingen voor die goedgekeurde tools. Een classificatie die uitsluitend naar de tool kijkt is onvolledig; hetzelfde model kan bij het ene gebruik onschuldig zijn en bij een ander gebruik aanzienlijk risicovoller, zodat het risico ontstaat op het snijvlak van tool en concreet gebruik, niet uit de tool alleen. Documenteer op kantoorniveau een register van goedgekeurde tools en toepassingen met risicoclassificatie en bijbehorende mitigerende maatregelen;

(3) Of, en onder welke omstandigheden, gebruik gemaakt mag worden door advocaten van gratis, publieke AI-tools. Ter voorkoming van misverstand - het onderscheid tussen 'gratis' en 'betaald' als maatstaf voor veiligheid is op zichzelf misleidend; er bestaan gratis, lokaal draaiende opensourcemodellen die volledig veilig zijn en dure commerciële tools die dat niet zijn.

Het relevante criterium is hoe de feitelijke verwerking is geregeld. Bepaal wat een advocatenkantoor wel en niet zelf kan en wil doen qua AI-automatisering. Wees daarbij kritisch en eerlijk over het eigen kennis- en expertiseniveau op het gebied van cybersecurity en applicatiebeheer. Open source en zelf gehoste oplossingen ogen aantrekkelijk vanwege de volledige controle, maar vergen doorlopend, deskundig beheer (e.g. patching, monitoring, toegangsbeheer). Zonder die deskundigheid kan een zelf gehoste oplossing in de praktijk onveiliger zijn dan een goed gecontracteerde, professioneel beheerde clouddienst.

(4) Dat de behandelend advocaat te allen tijde de volledige professionele en tuchtrechtelijke verantwoordelijkheid draagt, met een voorgeschreven werkwijze voor handmatige verificatie van AI-output;

(5) Een expliciete keuze over de mogelijkheid voor een cliënt tot AI opt-out; wordt een verzoek om AI niet te gebruiken in een bepaald dossier of voor een bepaalde cliënt gehonoreerd met het risico van hogere declaraties door handmatig werk, of geldt AI-ondersteuning als integraal onderdeel van de standaarddienstverlening waardoor opdrachten van weigerende cliënten niet worden aanvaard? Verduidelijkt moet worden hoe het kantoor omgaat met tijdschrijven, efficiëntiewinsten en eventuele doorbelasting van software- en licentiekosten, zodat discussies hierover bij de intake al zijn ondervangen in plaats van achteraf te ontstaan; en

(6) Verplichtingen tot periodieke scholing in AI-vaardigheden, prompting, bias en beveiliging, en het incidentprotocol (bijvoorbeeld voor de abusievelijke invoer van vertrouwelijke stukken in een ongeautoriseerde tool).

(7) De wijze van risico-inschatting moet blijken uit het AI-beleid. Overwogen kan worden om het risicomanagementproces aan te laten sluiten op industriestandaarden, zoals ISO-42001 met ISO-23894 of het NIST AI Risk Management Framework.

(8) Houd er rekening mee dat AI-functionaliteit niet altijd zichtbaar of bewust ingezet wordt: reguliere kantoorsoftware (zoals Word, Outlook of andere Microsoft- of Google-toepassingen) kan via automatische updates nieuwe of uitgebreide AI-functies toevoegen, en deze kunnen standaard geactiveerd zijn zonder dat de gebruiker daar expliciet voor heeft gekozen. Een whitelist van goedgekeurde tools is onvolledig als zij alleen ziet op de eerste ingebruikname en niet op latere functiewijzigingen van reeds gebruikte, ogenschijnlijk ‘niet-AI’ software. Het AI-beleid behoort daarom ook een proces te bevatten voor het signaleren en beoordelen van AI-functionaliteit die via updates van bestaande software wordt toegevoegd of geactiveerd, zodat deze vóór gebruik met zaakgegevens alsnog aan het beleid wordt getoetst.

Monitor de werking van het beleid structureel. Bij voorkeur aan de hand van geautomatiseerde (technische maatregelen), zoals netwerkmonitoring, maar ook aan de hand van organisatorische maatregelen, zoals voortdurende training en het afdwingen van personeelsbeleid. Herzie het beleid niet alleen volgens een vaste cyclus (bijvoorbeeld de jaarlijkse review), maar ook tussentijds zodra een wijziging bij een bestaande leverancier, een wijziging in wet- en regelgeving of een gemeld incident daartoe aanleiding geeft.

Wijs een AI-verantwoordelijke aan (bij grotere kantoren bijvoorbeeld de managing partner, CIO, CISO of Head of IT) die het beleid en de tools goedkeurt, verwerkersovereenkomsten onderhandelt, de jaarlijkse review uitvoert en escalatiepunt is bij incidenten. De advocaat zelf blijft eindverantwoordelijk voor alle output in een individueel dossier en beoordeelt de risico’s van concreet gebruik van AI in een specifieke zaak.

A.2 Selecteer, contracteer en beheer AI-leveranciers zorgvuldig

1 Aanbeveling

Bespreek vooraf met leveranciers hoe kwaliteit en waarborgen zijn geregeld, en doorloop bij de keuze van een AI-tool vier stappen: eerst oriënteren, dan een demo, dan uitproberen in een pilot, en pas daarna in gebruik nemen. Ga pas door naar de volgende stap als de vorige stap geen reden geeft om van de tool af te zien. Controleer vóór ingebruikname wie eigenaar is van de gegevens (i.e. prompts, contextuele gegevens en metadata), wie aansprakelijk is bij problemen, en hoe het contract weer kan worden beëindigd. Leg alle afspraken schriftelijk vast in een contract met de leverancier met heldere, eenduidige begrippen zodat het contract niet voor meerdere uitleg vatbaar is. Geloof de leverancier niet alleen op diens woord, maar test de belangrijkste beloften zelf tijdens een demo of proefperiode. Doe dit uitgebreider naarmate de tool met gevoeligere zaakgegevens gaat werken, en lichter bij eenmalig, beperkt gebruik of gebruik met uitsluitend openbare informatie. Overweeg bij risicovoller of complexer gebruik om een security audit naar de tool te laten uitvoeren (of de leverancier te vragen om de resultaten van een recente, onafhankelijke audit).

2 Toelichting

Deze aanbeveling is nodig omdat de keuze voor een AI-leverancier directe gevolgen kan hebben voor de naleving van alle kernwaarden van de advocatuur. Zonder een gestructureerde beoordeling loopt het kantoor het risico af te gaan op marketingbeloften of een FAQ-pagina in plaats van op daadwerkelijke, afdwingbare contractuele waarborgen. Een leverancier kiezen zonder vooraf na te denken over de exit leidt, bijvoorbeeld, tot ongewenste afhankelijkheid en het ontbreken van exit-afspraken wordt in de praktijk vaak pas gemerkt op het moment dat het kantoor wil overstappen en dan niet over de eigen gegevens kan beschikken; precies het moment waarop dit het kantoor het meest kwetsbaar maakt. De voorgestelde gefaseerde aanpak (i.e. oriënteren, demo, pilot, ingebruikname) is bedoeld om te voorkomen dat tijd en geld worden geïnvesteerd in een tool die al bij een eerste, lichte toets had moeten afvallen en het gebruik van eenduidige, heldere begrippen in het contract voorkomt dat de overeenkomst achteraf voor meerdere uitleg vatbaar blijkt.

Het is heel begrijpelijk dat advocaten zullen worstelen met het zelfstandig beoordelen van de geschiktheid van een IT-leverancier. Het inschakelen van gespecialiseerde hulp bij de product- en contractbeoordeling is daarbij toegestaan en raadzaam. Gelet op de kosten van IT-inkoop en -beheer wordt het – zeker kleinere – kantoren aangeraden om koopkracht te bundelen en kosten te verdelen door collectieve IT-inkoop en -beheerorganisaties op te richten en te benutten. Deze kunnen bijvoorbeeld worden opgezet binnen een algemene regionale context (e.g. speciaal voor advocatuur in regio Noord-Holland) of landelijk met het oog op een bepaalde specifieke deskundigheid (e.g. speciaal voor sociale advocatuur). Hieronder een overzicht van de aanbevolen minimumvereisten en gezichtspunten die advocaten moeten laten meewegen in hun risicobeoordeling en daarmee samenhangende besluitvorming over de inkoop en het gebruik van AI.

Minimumvereisten

(1) **Beschrijving van de dienst en onderaanneming.** Een heldere beschrijving van voorwerp en duur van de verwerking, een volledige lijst onderaannemers met notificatieplicht bij wijziging en of en hoe de dienst gebruikmaakt van een onderliggende modelleverancier of een tussenliggende middleware-laag, inclusief waar deze is gevestigd en welke waarborgen daar gelden.

(2) **Locatie van leverancier, dataopslag en dataverwerking.** Bij de keuze voor een leverancier, moet locatie meegewogen worden. Relevante gezichtspunten om mee te wegen bij zijn beoordeling zijn in ieder geval de locatie waar gegevens opgeslagen worden, maar ook waar de gegevens overigens verwerkt worden (e.g. inferentie, aanvullende functionaliteit, etc.), of de leverancier afhankelijk is van dienstverleners in andere landen, of de dienstverlener een niet-Europese moedermaatschappij heeft, of de gegevens versleuteld zijn en wie/waar de cryptografische sleutels worden beheerd. De aanbeveling is dat Nederlandse advocaten nastreven om de locatie van opslag en overige verwerking van geheimhouderinformatie technisch en juridisch zoveel mogelijk te beperken tot de Europese Economische Ruimte (EER). Dat vereenvoudigt naleving van gegevensbeschermingswet- en regelgeving (e.g. AVG). Daarnaast wordt verwacht dat Nederlandse advocaten ook beter bekend zijn met de mogelijkheden voor derden (waaronder civiele partijen, maar ook overheden, politie en justitie, inlichtingendiensten, etc.) om leveranciers te dwingen toegang te verlenen tot gegevens van advocaten, en ook met de waarborgen die advocaten kunnen inbouwen om hun cliënten hiertegen te beschermen. Houd er rekening mee dat gegevensopslag binnen de EER op zichzelf geen garantie biedt tegen toegang door buitenlandse overheden. Onder de Amerikaanse CLOUD Act, bijvoorbeeld, kan een leverancier - ook binnen de EER - verplicht zijn gegevens af te geven aan Amerikaanse autoriteiten zodra deze voldoende aanknopingspunten met de Verenigde Staten heeft. De analyse of een bepaalde leverancier onder de US Cloud Act (of vergelijkbare extraterritoriale wetgeving) valt, is leverancier-specifiek. Maar in algemene zin kan voor wat betreft de Cloud Act gesteld worden dat het risico pas wordt uitgesloten als de leverancier geen groepsverbondenheid heeft met een in de VS gevestigde onderneming en zelf ook onvoldoende aanknopingspunten met de VS heeft om Amerikaanse jurisdictie te rechtvaardigen (zoals het niet primair aanbieden van producten of diensten aan Amerikaanse klanten). Heeft de leverancier wel een dergelijke groepsverbondenheid, dan is het wenselijk dat de Amerikaanse groepsvennootschap geen bezit, bewaring of zeggenschap heeft over de in de EER opgeslagen gegevens. Een leverancier met een Amerikaanse moedermaatschappij zal hier niet snel aan voldoen, omdat de moedermaatschappij wordt geacht zeggenschap te hebben over de gegevens van haar dochterondernemingen. Waar gegevens buiten de Europese Economische Ruimte opgeslagen en/of verwerkt worden, maakt de advocaat vooraf een impactassessment om te beoordelen of dit verantwoord kan plaatsvinden. Het impactassessment toetst aan de kernwaarden, het eigen AI-beleid, en aan toepasselijke wet- en regelgeving (waaronder het zorgen voor toereikende waarborgen voor de doorgifte onder de AVG). In alle gevallen worden passende technische en organisatorische maatregelen getroffen door de advocaat om eventuele risico's zoveel mogelijk te mitigeren, waaronder die genoemd worden in de volgende paragraaf. Het eventuele restrisico en het besluit om dit te accepteren worden gedocumenteerd en periodiek geëvalueerd. Let er ook op dat afspraken worden gemaakt over 'geruisloze terugval' op een niet-goedgekeurde gegevenslocatie, zoals dat de leverancier zonder toestemming van de advocaat – bij storing van de primaire opslag- en verwerkingslocatie in, bijvoorbeeld, Frankfurt – terugvalt op een opslag- en verwerkingslocatie in Texas.

(3) **Vertrouwelijkheid en gegevensbescherming.** Beveiligingsmaatregelen zoals encryptie (in rust en in transport), eigen beheer van cryptografische sleutels, strenge toegangscontrole en periodieke penetratietesten (vraag om een recent pentest rapport in plaats van alleen een certificaat), gegevensisolatie (apart bewaren van gegevens ten opzichte van andere klanten), heldere afspraken of en onder welke voorwaarden prompts, context en output door werknemers kunnen worden ingezien bij de leverancier of een onderaannemer, en concrete toezeggingen vanuit de leverancier over wie (welke rollen) onder welke omstandigheden toegang heeft tot opgeslagen gegevens. In beginsel zouden medewerkers en onderaannemers van de leverancier geen onversleutelde zaakgegevens moeten kunnen inzien voor kwaliteitsborging of anderszins, tenzij het advocatenkantoor hiervoor toestemming heeft gegeven. Zie in dit kader verder aanbeveling C.2.

(4) **Zero Data Retention (ZDR).** ZDR is een contractuele en technische toezegging van een AI-leverancier dat ingevoerde prompts, context, overige input en gegenereerde output uitsluitend tijdelijk in het werkgeheugen (RAM/vRAM) van de computers van de leverancier worden verwerkt en direct na verzocht gebruik worden gewist, zodat niets wordt weggeschreven naar permanente schijfopslag, monitoringlogs of databases. Dat betekent ook geen gebruik van de gegevens voor pre-training, finetuning of Reinforcement Learning from Human Feedback (RLHF) en geen serverside caching van prompt-contexten of documentfragmenten over sessies heen. Vraag de leverancier dit expliciet te bevestigen voor zowel de eigen systemen als die van een onderliggende modelleverancier of tussenliggende middleware-laag, en laat dit zo mogelijk verifiëren tijdens de pilot en verifieer periodiek of de leverancier zich hieraan houdt. Voor zover de functionaliteit van chatgeschiedenis nodig is en in de weg staat aan ZDR, worden afspraken gemaakt over het geautomatiseerd en onherroepelijk verwijderen van dergelijke gegevens op verzoek.

(5) **Toegang, herstel en teruggave van gegevens bij einde dienstverlening.** Verwijdering van gegevens binnen een bepaalde periode (e.g. 10 werkdagen) na contractbeëindiging, inclusief een duidelijke afspraak over wat er gebeurt met technisch afgeleide gegevens (embeddings) van documenten, en data-export in een machine-readable, algemeen gebruikt formaat. Zie in dit kader verder aanbeveling A.3.

(6) **Onderhoud en wijzigingen.** Een beschrijving van afspraken over welk onderhoud wordt uitgevoerd, door wie, hoe vaak, en bij voorkeur buiten kantooruren, en een waarborg dat de leverancier contractvoorwaarden niet eenzijdig en op korte termijn kan wijzigen zonder voorafgaand overleg.

(7) **Bijstand bij een incident.** Vastgelegde reactie- en oplostijden bij een gemelde storing of incident, de beschikbaarheid van ondersteuning (bijvoorbeeld of er 24x7-service is), en een omschrijving van de rol en verplichtingen van de leverancier bij een cyberincident, aansluitend bij het eigen incidentresponsplan van het kantoor.

(8) **Medewerking aan toezicht.** Ten minste eenmaal jaarlijks uit te oefenen auditrechten, aangevuld met de verplichting van de leverancier om mee te werken aan security audits naar aanleiding van een incident of redelijk vermoeden van inbreuk op wet- en regelgeving of het contract, en regelmatige rapportage door de leverancier.

(9) **Aansprakelijkheid en verzekering.** Een toereikende aansprakelijkheidsregeling voor wanprestatie van de leverancier, zonder onredelijke beperking van de aansprakelijkheid (in het bijzonder bij een datalek of ander ernstig incident), en bevestiging dat de leverancier voor deze aansprakelijkheid adequaat verzekerd is.

(10) **Rechten ten aanzien van gegevens.** Advocaten behouden en waar nodig verkrijgen van de leverancier alle IE-rechten en overige rechten op hun prompts, context, geüploade bestanden en overige invoer, maar ook de door AI gegenereerde output en andere gegevens die gegenereerd of verkregen worden als gevolg van het gebruik door de advocaat. Leverancier moet ervoor instaan dat deze toezeggingen ook gelden voor onderaannemers waar de leverancier van gebruikmaakt (e.g. onderliggende LLM-aanbieders).

(11) **Beëindiging en minimale opzegtermijnen.** Een redelijke opzegtermijn zonder excessieve beëindigingsvergoedingen.

(12) **Volledige, meetbare dienstverleningsniveaus.** Concrete afspraken over de beschikbaarheid (uptime) van de systemen, waarbij een gegarandeerd percentage in de praktijk wordt doorgerekend (bij 99,9% per jaar mag een systeem een ruime werkdag per jaar uitvallen, bij 99,9% per maand minder dan een uur), met een afspraak over hoe lang de dienst offline mag zijn voor onderhoud of storingen en of dit binnen de uptime of de downtime valt.

(13) **Testen van bedrijfscontinuïteit en beveiliging.** Eigen, actieve verificatie van kritieke waarborgen en toezeggingen tijdens een pilot in plaats van alleen op de verklaringen van de leverancier te vertrouwen, onder meer op toegangsscheiding tussen gebruikers, daadwerkelijke verwijdering van gegevens, en weerbaarheid tegen prompt-injectie, aangevuld met een gezamenlijk herstelplan voor de terugkeer naar normale bedrijfsvoering na een incident.

(14) **Doorlopend toezicht op de prestaties van de leverancier.** Periodieke controles of de beveiligingsmaatregelen en het beleid van de leverancier daadwerkelijk werken, en het recht om bij vastgestelde tekortkomingen aanvullende, tijdgebonden afspraken te maken of corrigerende maatregelen te eisen.

(15) **Uitgewerkte exitstrategie met overgangsperiode.** Een exitplan dat al vóór het aangaan van de overeenkomst is opgesteld, met afspraken over medewerking, continuïteit van de dienstverlening en verantwoordelijkheden tijdens de exit, de overdracht van alle relevante gegevens inclusief back-ups, en afspraken over de wijze van overdracht en de vernietiging van gegevens na een succesvol afgeronde overstap.

(16) **AI-geletterdheid van de leverancier.** Een toezegging dat de leverancier zijn eigen personeel (en dat van onderaannemers) dat betrokken is bij het leveren van de AI-tool adequaat getraind houdt op verantwoord gebruik van AI, en dat de leverancier het kantoor de documentatie verschaft die nodig is om op zijn beurt de eigen AI-geletterdheid van medewerkers te waarborgen.

(17) **Verbod op en bewaking van verboden AI-praktijken.** Een garantie van de leverancier dat de AI-tool voorafgaand aan levering is getest en doorlopend wordt gemonitord om te waarborgen dat deze geen praktijken gebruikt die wettelijk verboden zijn, met een meldplicht en een korte hersteltermijn als dit alsnog wordt geconstateerd.

(18) **Opzegging bij overdracht van de overeenkomst.** Een recht om de overeenkomst op te zeggen bij contractsovername als de overnemende derde partij niet redelijkerwijs in staat geacht kan worden de overeenkomst met minimaal dezelfde kwaliteit (waaronder nakoming van deze minimumvereisten) voort te kunnen zetten of als de overnemende derde partij gevestigd is in een land buiten de Europese Economische Ruimte.

A.3 Houd rekening met afhankelijkheid van een beperkt aantal leveranciers en dataportabiliteit

1 Aanbeveling

Het kantoor betrekt uitdrukkelijk het risico van vendor lock-in (i.e. de afhankelijkheid van een of meerdere technologieleveranciers) expliciet in de besluitvorming over de inzet van AI-tools bij kernprocessen en maakt passende afspraken rondom dataportabiliteit.

2 Toelichting

IT in algemene zin, maar AI in het bijzonder, lijkt een essentieel hulpmiddel voor onze beroepsgroep te zullen vormen. Het kan de kernwaarde onafhankelijkheid raken wanneer kernprocessen van het kantoor, of van de beroepsgroep als geheel, afhankelijk worden van één of een kleine groep aanbieders. Aanbevolen is om het risico van langdurige of permanente onbeschikbaarheid (of niet-acceptabele prijsstijging) van een aanbieder of AI-systeem op passende wijze te waarborgen. Maak ook passende afspraken over de exportmogelijkheden van kantoor- en cliëntgegevens om het overstappen van één aanbieder naar een andere te vergemakkelijken. Deze aanbeveling weegt zwaarder naarmate een tool een groter deel van de kernprocessen van het kantoor ondersteunt; voor incidenteel of ondersteunend gebruik is een lichte toets voldoende.

Indien nodig blijkt om over te stappen naar een andere leverancier, is het wenselijk dat de advocaat met de leverancier heeft afgesproken dat er sprake is van technische voorzieningen die de advocaat de mogelijkheid geven om zijn data te exporteren en te hergebruiken – al dan niet na conversie – bij een nieuwe leverancier. Het gaat hierbij in ieder geval om de volgende data:

- **Geüploade bestanden.** Alle processtukken, adviezen, dossiers, correspondentie en overige elektronische bestanden in hun oorspronkelijke vorm
- **Prompts, promptgeschiedenis en metadata.** Volledige promptgeschiedenis met bijbehorende gegenereerde output en eventuele daaraan gerelateerde gebruiksgegevens.
- **Prompt libraries.** Sjablonen of standaardprompts die door het kantoor of de advocaat zijn opgesteld.
- Indien van toepassing, technisch mogelijk voor de bestaande leverancier en voor zover bruikbaar voor de nieuwe leverancier, **kantoorspecifieke adapter-bestanden en modelgewichten** die gegenereerd kunnen zijn naar aanleiding van maatwerk of finetuning op eerdergenoemde categorieën van data. In alle gevallen dient de bestaande leverancier toe te zeggen deze categorie data niet te gebruiken voor enig ander doel dan de dienstverlening aan het desbetreffend advocatenkantoor en deze data te verwijderen binnen een bepaalde en redelijke tijd na beëindiging van de overeenkomst.

Gegevens moeten zoveel mogelijk worden geleverd in gestructureerde industriestandaarden, zoals .json, .csv of .xml, maar met uitzondering van geüploade bestanden die in hun oorspronkelijke bestandsformat verstrekt moeten worden.

A.4 Informeer leveranciers en cliënten over geheimhouderinformatie

1 Aanbeveling

Wanneer de advocaat verwacht dat een leverancier toegang zal krijgen tot geheimhouderinformatie, informeert de advocaat de leverancier daar voorafgaand uitdrukkelijk en schriftelijk over en spreekt de advocaat schriftelijk een protocol af dat beschrijft hoe de leverancier moet omgaan met verzoeken van derden (e.g. private procespartijen, politie en justitie, inlichtingendiensten, toezichthouders) om dergelijke gegevens.

De advocaat informeert bovendien ook diens cliënten (bijvoorbeeld, door middel van de opdrachtbevestiging) dat wanneer de cliënt zelf correspondentie met, of stukken van of voor, de advocaat invoert in publieke AI-tools (zoals ChatGPT), dit de vertrouwelijkheid van die informatie aantast en in voorkomend geval zelfs de geheimhouderinformatie-status kan doen vervallen.

2 Toelichting

AI-leveranciers (en IT-leveranciers in bredere zin, zoals hosting- en cloudopslagdiensten) worden een steeds onmisbaarder onderdeel van de bedrijfsvoering van advocatenkantoren, onder meer voor dossieranalyse, onderzoek, documentgeneratie en communicatie. Daardoor krijgen zij in toenemende mate geheimhouderinformatie onder zich — informatie die naar haar aard onder de geheimhoudingsplicht van advocaten valt en ten aanzien waarvan advocaten verschoningsgerechtigd zijn. Die informatie loopt daardoor het risico om te worden opgevraagd door een derde. Zodra het redelijk vermoeden bestaat dat zich bij gevorderde gegevens geheimhouderinformatie bevindt, moet de advocaat in beginsel in staat worden gesteld zich uit te laten over diens verschoningsrecht met betrekking tot die informatie. Dat is pas mogelijk als de advocaat tijdig op de hoogte wordt gebracht van een eventueel verzoek om dergelijke informatie. Om die reden wordt advocaten aanbevolen om de volgende afspraken schriftelijk vast te leggen met hun leverancier. Dit is mede van belang omdat bij de beoordeling van een mogelijk vormverzuim en de daaraan te verbinden gevolgen relevant kan zijn in hoeverre de advocaat zelf passende maatregelen heeft genomen om diens informatiehuishouding zo in te richten dat de kans op inbreuken op het verschoningsrecht zo gering mogelijk blijft.

- **Geen rechtstreekse verstrekking.** De leverancier verstrekt geen data rechtstreeks aan een derde zonder voorafgaande schriftelijke toestemming van de advocaat, tenzij de leverancier hier dwingendrechtelijk toe verplicht is. De leverancier verwijst zoveel mogelijk de verzoekende partij in schriftelijk en onverwijld door naar het advocatenkantoor, met expliciete verwijzing naar het beroepsgeheim van de advocaat.

- **Voorafgaande en onverwijlde melding aan de advocaat.** De leverancier stelt de advocaat onverwijld schriftelijk op de hoogte van het bestaan, de inhoud en de reikwijdte van het verzoek, voordat enige data wordt verstrekt, zodat de advocaat de gelegenheid heeft de gegevens te beoordelen voordat een derde daar kennis van kan nemen. Indien een geheimhoudingsbevel dit verbiedt, maakt de leverancier daartegen onverwijld bezwaar en vecht de leverancier dat geheimhoudingsbevel zo nodig in rechte aan op kosten van het advocatenkantoor. Over de kostenvergoeding daarvoor maakt de advocaat vooraf afspraken met de leverancier.

- **Verweer en uitputting van rechtsmiddelen.** Wanneer het verzoek niet kan of mag worden gedeeld met de advocaat, wendt de leverancier alle redelijke rechtsmiddelen aan (bezwaar, beroep, kort geding of een vergelijkbare voorlopige voorziening) op kosten van het advocatenkantoor tegen een verzoek dat het beroepsgeheim mogelijk schendt, onnodig ruim of onevenredig is, of de toepasselijke wettelijke waarborgen lijkt te omzeilen. Over de kostenvergoeding daarvoor maakt de advocaat vooraf afspraken met de leverancier.

- **Minimale verstrekking als uiterste middel.** Voor zover verstrekking noodzakelijk blijft, verstrekt de leverancier uitsluitend het strikt noodzakelijke, in de vorm waarin de data zich bevindt (versleuteld, indien van toepassing). De leverancier informeert de advocaat na verstrekking onverwijld over de omvang en inhoud van wat is verstrekt.

Risico vanuit de cliënt zelf

Naast het risico dat ontstaat doordat de advocaat zelf een AI gebruikt, doet zich een bijzonder risico voor wanneer de cliënt buiten medeweten van de advocaat zelf correspondentie met of stukken van de advocaat invoert in een publieke AI-tool, zoals ChatGPT. Een dergelijke publieke tool slaat vaak zowel de ingevoerde als de gegenereerde informatie op en kan deze gebruiken voor onder meer het trainen van het publieke AI-model. Wanneer zo'n tool gebruikt wordt om tekst te verwerken die bestemd is voor, of afkomstig is van, een advocaat, kan dit worden aangemerkt als het door de cliënt openbaar maken van die informatie. Daardoor kan die informatie mogelijk haar karakter als geheimhouderinformatie verliezen, met mogelijk ernstige gevolgen voor de cliënt, naast de algemene zorgen over informatiebeveiliging bij gebruik van dergelijke publieke AI. Dit probleem kan zich ook voordoen als de cliënt zelf bij een vertrouwelijk gesprek met de advocaat gebruik maakt van een publieke AI-transcriptietool. De cliënt is zich hiervan vaak niet bewust, en actieve instructie door de advocaat kan helpen dit risico te beperken.

B Bekwaamheid en gebruik

B.1 Investeer in kennis over (generatieve) AI

1 Aanbeveling

Iedere advocaat verwerft en onderhoudt basiskennis van (generatieve) AI, waaronder de werking van taalmodellen, prompt-vaardigheid, foutdetectie, bias-mitigatie, cybersecurity en relevante regelgeving.

2 Toelichting

Basiskennis (of AI-geletterdheid) is een voorwaarde voor verantwoorde inzet en voor het herkennen van risico's zoals hallucinaties en onjuiste bronvermelding; dit raakt primair deskundigheid. Deze norm vereist niet dat elke advocaat diepgaande technische achtergrondkennis heeft van de onderliggende architectuur van taalmodellen; het gaat om functioneel begrip dat nodig is om AI-output te kunnen beoordelen en risico's te herkennen. Het vereiste kennisniveau is voor elke advocaat gelijk voor zover het gaat om risicobewustzijn (bijvoorbeeld: dat output foutief kan zijn, dat vertrouwelijke informatie aan derden kan worden ontsloten). Verdergaande kennis, zoals geavanceerde prompt-technieken of inzicht in retrievaltechnieken, is met name relevant voor advocaten die AI bovengemiddeld intensief of voor complexere toepassingen gebruiken. Denk hierbij aan het volgen van opleidingen over (generatieve) AI, promptengineering, bias-mitigatie, hallucinaties, foutdetectie, juridisch relevante regelgeving (waaronder de AI-verordening) en cybersecurity. Houd kennis actueel via permanente educatie en het intern delen van best practices, bijvoorbeeld via gestructureerde feedbackmomenten. Besteed in de beroepsopleiding expliciet aandacht aan hoe een taalmodel met externe bronnen kan worden verrijkt (bijvoorbeeld via technieken die het model verbinden met actuele wetgeving en jurisprudentie), zodat betrouwbaarder resultaten kunnen worden verkregen dan met een kaal taalmodel. Documenteer gevolgde opleidingen per medewerker als onderdeel van de reguliere permanente-educatieregistratie.

B.2 Doe praktische ervaring op met AI-tools

1 Aanbeveling

In aanvulling op het uitbreiden van hun theoretische kennis over technologie en AI, moeten advocaten actief praktijkervaring opdoen met het gebruik van AI-tools in overeenstemming met hun kantoorbeleid.

2 Toelichting

Praktische ervaring is de meest effectieve manier om intuïtie te ontwikkelen voor de mogelijkheden en grenzen van AI; dit is niet via een geschreven document over te dragen.

Deze norm vereist niet dat elke advocaat op eigen initiatief met ongeautoriseerde tools experimenteert: het experimenteren vindt plaats binnen de kaders van het AI-beleid. Deze aanbeveling geldt in gelijke mate voor iedere advocaat, ongeacht rechtsgebied of ervaringsniveau. Faciliteer als kantoor gestructureerde experimenteerruimte, bijvoorbeeld via interne workshops of 'vlieguren'-programma's, zodat kennisopbouw niet uitsluitend van individueel initiatief afhangt.

B.3 Verifieer altijd het eindresultaat alvorens het te delen met procespartijen en rechtspraak

1 Aanbeveling

De advocaat controleert steeds AI-gegenereerde output op juistheid en volledigheid voordat deze wordt gebruikt in advies, correspondentie of processtukken. De intensiviteit van de controle moet proportioneel zijn aan de mogelijke risico's van eventuele fouten. Afwijken van dit uitgangspunt moet mogelijk zijn met toestemming van de cliënt nadat de advocaat de cliënt heeft geïnformeerd over mogelijke risico's.

2 Toelichting

AI-tools kunnen citaten, jurisprudentie of feiten verkeerd weergeven. De advocaat blijft – net als bij het accepteren van concepten van stagiairs - verantwoordelijk voor de juistheid van alles wat uiteindelijk aan de cliënt, procespartijen of rechter wordt voorgelegd. Het bewust of nalatig overnemen van hallucinaties kan in strijd zijn met de waarheidsplicht van artikel 21 Rv. Weeg de verificatie-intensiteit af tegen de mogelijke impact van fouten; bij onherstelbare gevolgen, zoals een verkeerd cassatieadvies op basis van gehallucineerde rechtspraak, is de impact groter. Verifieer daarom jurisprudentie via het ECLI-nummer en de originele overweging (niet alleen de AI-samenvatting) bij een officiële bron en controleer of de uitspraak bestaat, relevant is, en of deze niet inmiddels achterhaald is door recentere rechtspraak. Verifieer het bestaan en de inhoud van wetsartikelen via officiële publicaties en let in het bijzonder op data van inwerkingtreding, overgangsrecht en artikelnummering. Controleer bij feitelijke stellingen of ze op de juiste manier gestaafd worden door het dossier. Documenteer wanneer afgeweken wordt van het uitgangspunt dat eindresultaten handmatig door de advocaat worden gecontroleerd. Bij bulkwerk mag, bijvoorbeeld, een gedocumenteerde steekproef met toestemming van cliënt volledige handmatige controle vervangen, mits de output op een representatieve subset is getest en betrouwbaar bevonden.

B.4 Wees alert op sycophancy

1 Aanbeveling

De advocaat is alert op de neiging van AI-systemen om antwoorden te genereren die aansluiten bij de kennelijke voorkeuren van de gebruiker (sycophancy), ook wanneer de onderliggende aanname of vraagstelling onjuist is.

2 Toelichting

Sycophancy is een inherent kenmerk van taalmodellen in het algemeen, ongeacht de leverancier; het risico zit niet in een specifieke prompt maar in de neiging van het model. Dit raakt primair onafhankelijkheid: de advocaat moet voorkomen dat diens eigen (voorlopige) oordeel ongemerkt wordt bevestigd in plaats van getoetst. Deze norm impliceert niet dat bepaalde tools per definitie minder gevoelig zijn voor sycophancy dan andere. Deze aanbeveling weegt zwaarder naarmate de advocaat AI inzet om een al ingenomen standpunt te toetsen of te onderbouwen, en minder zwaar bij zuiver informatieve of administratieve toepassingen. Vraag het model, bijvoorbeeld, expliciet om tegenargumenten en laat het model aannames betwisten in plaats van bevestigen. Gebruik, bijvoorbeeld, bewust tegengestelde prompts om de robuustheid van een conclusie te testen, en stel jezelf actief de vraag of de AI-output een alternatieve lezing heeft overwogen. Overweeg het gebruik van vaste systeeminstructies op dossier- of kantoorniveau die objectiviteit en brongrondslag afdwingen en het risico op sycophancy beperken.

B.5 Wees alert bij het gebruik van agentic AI

1 Aanbeveling

De advocaat stelt in het AI-beleid (of informatiebeveiligingsbeleid) expliciete grenzen aan de autonomie van AI-agents, zorgt voor menselijk toezicht op elke stap die rechtsgevolgen kan hebben, en gebruikt geen agentic AI zonder dat de werkwijze, de aangeroepen tools en de datastroom volledig in kaart zijn gebracht en de risico's daarvan zijn afgewogen en in overeenstemming zijn bevonden met het voornoemd beleid.

2 Toelichting

Agentic AI-systemen plannen en voeren zelfstandig taken uit, roepen externe tools aan en kunnen uitgebreide reeksen aan subtaken uitvoeren zonder voortdurende menselijke bevestiging. De risico's verschillen wezenlijk van die van simpele chatbots: agenten handelen daadwerkelijk (bijvoorbeeld door e-mails te versturen of documenten te wijzigen), fouten kunnen zich voortplanten doorheen de keten van systemen waartoe de agents toegang hebben, en agenten die externe bronnen raadplegen zijn kwetsbaar voor misleidende instructies daarin. Dit raakt de onafhankelijkheid en deskundigheid.

Deze aanbeveling sluit het gebruik van agentic AI niet uit, maar koppelt het gebruik van deze technologie aan een voorafgaande inventarisatie en aan doorlopend menselijk toezicht. De mate van menselijk toezicht is afhankelijk van de reikwijdte van het mandaat dat aan de agent wordt gegeven en van de vraag of de uitkomst rechtsgevolgen kan hebben voor de cliënt of derden. Bij een beperkte, goed afgebakende taak (bijvoorbeeld het ophalen van openbare bronnen) volstaat controle achteraf; bij taken met directe uitwerking naar de cliënt of de rechtbank (bijvoorbeeld het verzenden van correspondentie) is intensiever toezicht wenselijk. Breng vóór gebruik in kaart welke externe API's, databases of diensten van derden de agent kan aanroepen, en welke vertrouwelijke gegevens daarbij langs meerdere verwerkingsverantwoordelijken kunnen stromen. Dwing beperkingen af met technische maatregelen (e.g. op firewall-niveau). Zorg voor traceerbaarheid en logging van de handelingen van de agent, mede met het oog op eventuele beroepsaansprakelijkheid, en voor controle achteraf voordat een resultaat aan de cliënt of de rechtbank wordt vrijgegeven. Beoordeel vooraf ook het risico van mandaatinterpretatie: de mogelijkheid dat een agent instructies breder uitlegt dan bedoeld.

B.6 De advocaat blijft eindverantwoordelijk; gebruik AI als hulpmiddel

1 Aanbeveling

De advocaat blijft eindverantwoordelijk voor diens correspondentie, processtukken of advisering. AI kan in grote of kleine mate worden ingezet, maar het blijft een hulpmiddel en niet een vervanging van de eigen juridische beoordeling en eindverantwoordelijkheid. Het is belangrijk dat voordat de door de AI gegenereerde output wordt gedeeld met derden, de advocaat de output zelf begrijpt, onderschrijft en heeft gecontroleerd.

2 Toelichting

Dit raakt de kern van de kernwaarde onafhankelijkheid en deskundigheid. 'Hulpmiddel' sluit intensief gebruik niet uit, zolang de advocaat de uiteindelijke beoordeling zelf maakt en kan verantwoorden wat en waarom iets in een eindproduct zo is opgeschreven. Onafhankelijkheid staat niet alleen onder druk bij bewuste, herkenbare beïnvloeding zoals sycophancy; het risico is groter wanneer AI-output de beoordeling van de advocaat ongemerkt bijstuurt, zonder dat de advocaat zich daarvan bewust is. Dit onbewuste beïnvloedingsrisico neemt toe naarmate een advocaat vaker en met meer vertrouwen op dezelfde tool leunt. De mate van AI-inzet die verantwoord is, hangt af van het type taak: (a) een beperkte, procedurele deeltaak, bijvoorbeeld het opmaken van een productie-index of het formatteren van een document; (b) verbetering van het resultaat van een reeds voltooide menselijke beoordeling, bijvoorbeeld het laten verbeteren van spelling en leesbaarheid van een zelf geschreven concept; (c) het opsporen van patronen of afwijkingen zonder dat dit de eerdere menselijke beoordeling vervangt of beïnvloedt zonder toetsing, bijvoorbeeld het laten markeren van afwijkende clausules in honderd contracten die vervolgens één voor één zelf worden beoordeeld; of (d) een voorbereidende taak ten behoeve van een beoordeling die relevant is voor het advies, bijvoorbeeld een eerste juridische analyse van een dossier die de advocaat vervolgens zelf volledig opnieuw beoordeelt. Naarmate de AI-bijdrage dichter bij de uiteindelijke inhoudelijke beoordeling komt of die bijdrage gedeeld wordt met procespartijen of rechtspraak, is intensievere controle door de advocaat vereist. Wees ook alert op druk vanuit de cliënt zelf om blind te varen op een AI-analyse omwille van tijd- of kostenbesparing en bewaak hierin de eigen onafhankelijkheid.

B.7 Bewaak de vakbekwaamheidsontwikkeling van junioren

1 Aanbeveling

Het gebruik van AI mag niet leiden tot cognitieve atrofie binnen de advocatuur. Het kantoor zorgt ervoor dat junioradvocaten basisvaardigheden, zoals kritisch nadenken, juridisch onderzoek uitvoeren, juridisch-analytisch schrijven, handmatig blijven ontwikkelen - los van AI-ondersteuning. Tegelijk wordt de vaardigheid om AI effectief en verantwoord te gebruiken en de mogelijkheid om AI te mogen gebruiken naar verwachting een steeds zwaarder wegend vereiste bij de werving van jong talent; junioren zijn doorgaans beter op de hoogte van de mogelijkheden en beperkingen van de technologie zelf, maar missen vaak nog het juridisch-analytisch denkmodel om die vaardigheid in de advocatuur goed te kunnen toepassen.

2 Toelichting

Door het automatiseren van routinetaken dreigt de fase te verdwijnen waarin junioren hun juridische intuïtie ontwikkelen; het risico is een generatie die technisch vaardig is met AI, maar de diepgang mist om te herkennen wanneer AI de plank misslaat. Dit raakt primair deskundigheid, met een langetermijneffect op de kwaliteit van de gehele beroepsgroep. Deze norm verbiedt geenszins dat junioren AI gebruiken, maar vereist dat dit gebruik niet in de plaats komt van het zelfstandig en handmatig doorlopen van kernleermomenten. Verder geldt dat er situaties kunnen ontstaan waarin er geen AI beschikbaar is. In voorkomende gevallen dienen deze taken onverminderd uitgevoerd te kunnen worden. Deze aanbeveling vormt daarmee ook een oproep aan de beroepsgroep om nieuwe manieren te ontwikkelen – in samenwerking met de beroepsopleidingen en de universiteiten – om vakkennis over te dragen en te toetsen, aanvullend op werkvormen die door AI zijn geautomatiseerd. Dit onderwerp wordt daarnaast opgenomen in het reguliere opleidingsbeleid van het kantoor.

B.8 Stel een eigen testdataset samen en gebruik deze bij het periodiek evalueren van bestaande en nieuwe leveranciers

1 Aanbeveling

Stel een eigen (geanonimiseerde) testdataset samen van adviezen, processtukken, presentaties en correspondentie die volledig handmatig zijn opgesteld, gecontroleerd en akkoord zijn bevonden. Deze testdataset dient zoveel mogelijk representatief te zijn voor de belangrijkste werkzaamheden die het advocatenkantoor verricht. Houd deze testdataset actueel voor het objectief testen en valideren van bestaande en nieuwe AI-software.

2 Toelichting

Demonstraties van AI-aanbieders maken veelal gebruik van documenten en taken die optimaal aansluiten bij het trainingsmodel. De dagelijkse advocatenpraktijk kenmerkt zich echter door imperfecte gegevens, complexiteit en specifieke, afwijkende formats. Vertrouw bij de selectie en evaluatie van AI-tools daarom niet uitsluitend op demonstraties of vooraf geselecteerde testbestanden van leveranciers. Bouw als kantoor een representatieve testdataset (zoveel mogelijk geanonimiseerd) voor de belangrijkste werkzaamheden waarbij AI gebruikt zal worden. Gebruik deze dataset bij periodieke update van bestaande software en evaluatie van nieuwe software om AI-output te benchmarken tegen een menselijke kwaliteitsnorm. Naarmate de testdataset vaker gebruikt wordt, worden prestaties beter meetbaar en onderling vergelijkbaar. Actualiseer de testdataset periodiek met nieuwe jurisprudentie, wetswijzigingen en recente casuïstiek uit de kantoorpraktijk, zodat de testdataset relevant blijft.

C Informatiebeveiliging

C.1 Pas dataminimalisatie toe

1 Aanbeveling

Deel bij het gebruik van AI-tools uitsluitend redelijkerwijs noodzakelijke informatie. De advocaat is zich bewust van de mogelijke informatiebeveiligingsrisico's en deelt informatie daarom alleen doelmatig. Want informatie die niet gedeeld is met een derde kan in beginsel ook niet door die derde gelekt worden.

2 Toelichting

Deze aanbeveling sluit niet uit dat grote hoeveelheden data met behulp van AI worden verwerkt (bijvoorbeeld bij bulkwerk zoals due diligence of e-discovery), maar vereist wel dat de advocaat zich bewust is van de risico's van die verwerking en om die reden doelmatig is in de keuze wat wel en niet wordt gedeeld. Deze aanbeveling vereist op zichzelf geen categorische anonimisering of pseudonimisering van alle zaakgegevens voor zover de gebruikte tooling over adequate informatiebeveiligingswaarborgen beschikt. Dergelijke maatregelen zijn wel raadzaam als aanvullende waarborg bij hoog-risico gebruik (zie AI-beleid).

Deze norm raakt vertrouwelijkheid en volgt uit de algemene verplichting om zorgvuldig met cliëntgegevens om te gaan; weet daarom wat wordt ingekocht: welke data wordt gebruikt (bijvoorbeeld prompts en context), waar data wordt opgeslagen en waar deze daadwerkelijk wordt verwerkt - dataopslag en inferentie zijn niet noodzakelijk dezelfde locatie - en begrijp de technologie en de voorwaarden van de leverancier vóór inkoop en gebruik. Wees ervan bewust dat niet alleen invoergeschiedenis, maar ook metadata en gebruiksdata (zoals locatiegegevens of geopende bestanden) door de aanbieder kunnen worden verwerkt, naast de bewuste input. De aanbeveling vraagt dus niet om omvang of schaal van de verwerking op zich te beperken, maar om bewust te kiezen welke informatie voor het beoogde gebruiksdoel nodig is, zodat passende waarborgen voor beroepsgeheim en informatiebeveiliging het eigenlijke zwaartepunt blijven.

C.2 Zorg voor aantoonbare beveiliging en AI-specifieke beheersmaatregelen

1 Aanbeveling

Zorg voor aantoonbare informatiebeveiliging van AI-systemen over drie lagen: organisatiebeheersing, applicatiebeveiliging en AI-specifieke weerbaarheid. De advocaat legt contractueel de afspraken hierover vast en haakt hierbij, waar mogelijk, aan bij industriestandaarden.

2 Toelichting

Proces- en organisatiebeheersing
De basis van betrouwbare dienstverlening ligt in een volwassen risicomanagementsysteem en een veilige softwareontwikkelingscyclus. De advocaat eist van de leverancier contractueel vastgelegde processen en onafhankelijke toetsing op het gebied van:

- Structureel risicomanagement en beleid, zoals periodieke dreigings- en risicoanalyses gedurende de hele levenscyclus van de software.
- Scheiding van omgevingen en wijzigingsbeheer, zoals een strikte scheiding tussen ontwikkel-, test-, acceptatie- en productieomgevingen (OTAP) om ongecontroleerde wijzigingen uit te sluiten.
- Onafhankelijke toetsing en verificatie door middel van periodieke audits, kwetsbaarheidsscans en penetratietests (pentests).
- Operationeel beheer met vastgelegde processen voor incidentrespons, gecontroleerd patchbeheer en veilige logging/monitoring.

Voorbeelden van relevante standaarden:

- ISO/IEC 27001/2
- SOC 2 (Type II)
- NCSC ‘Beleids- en beheersingsrichtlijnen voor veilige software’

Basisbeveiliging van applicatie en infrastructuur
Vrijwel alle AI-toepassingen voor de advocatuur worden ontsloten via traditionele webapplicaties. Kwetsbaarheden in deze laag kunnen aanvallers directe toegang geven tot de achterliggende AI-systemen en modellen. Essentiële technische principes hierbij zijn:

- Strikte invoer- en uitvoerbehandeling. Invoer mag nooit impliciet worden vertrouwd en moet op de server worden genormaliseerd en gevalideerd vóór verwerking; dynamische uitvoer moet veilig gecodeerd worden om injectieaanvallen te voorkomen.

- Toegangsbeheer en identiteit (IAM). Toepassing van het least privilege-beginsel, robuuste rolgebaseerde autorisatie en sterke authenticatie (bij voorkeur phishingresistente multi-factor authenticatie).

- Encryptie. Versleuteling van data in rust (databases/opslag) en data in transit middels actuele transportbeveiliging (TLS/HTTPS).

- Systeemhardening en netwerkisolatie: Het minimaliseren van de aanvalsmogelijkheden door onnodige poorten/services uit te schakelen en kritieke backend-systemen af te schermen achter een gedemilitariseerde zone (DMZ).

Voorbeelden van standaarden:

- NCSC ICT-beveiligingsrichtlijnen voor webapplicaties

- OWASP Top 10 / ASVS (Application Security Verification Standard)

AI-specifieke dreigingen
AI kent unieke kwetsbaarheden die zich afspelen in de interactie met het AI-systeem en model zelf (e.g. training, context en inferentie). De leverancier moet daarom aantoonbaar maatregelen nemen tegen:

- Adversarial input en prompt manipulation. Aanvallen waarbij kwaadwillenden via gerichte prompts of gemanipuleerde processtukken proberen de ingebouwde veiligheidsregels van het model te omzeilen (jailbreaking) of het model onbedoelde acties te laten uitvoeren.

- Datalekken via modeluitvoer (Data Extraction / Inference). Het risico dat vertrouwelijke cliëntgegevens of dossiers via eerdere interacties, fine-tuning of contextvensters herleidbaar worden en weglekken naar niet-geautoriseerde gebruikers.

- Data- en contextvergiftiging (Data Poisoning). Manipulatie van bronbestanden binnen Retrieval-Augmented Generation (RAG) of opleidingsdata, wat kan leiden tot gemanipuleerde of onbetrouwbare juridische uitkomsten.

Voorbeelden van standaarden:

- ISO/IEC 42001 (voor het beheersysteem van AI) in combinatie met ISO/IEC 23894 (voor AI-risicomanagement)

- NIST AI Risk Management Framework

- MITRE ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems): Kennisbank en taxonomie van reële aanvalstechnieken specifiek gericht op AI- en machine learning-systemen.

- OWASP Top 10 for Large Language Models (LLMs): Praktisch overzicht van de tien meest voorkomende kwetsbaarheden in generatieve AI-systemen.

Certificeringen en auditrapporten
ISO-certificeringen en vergelijkbare auditrapporten kunnen een nuttig, maar nooit doorslaggevend, hulpmiddel zijn bij de beoordeling van een AI-leverancier. Het ontbreken van relevante certificering of auditrapporten kan meewegen bij de vraag of een leverancier in overeenstemming handelt met industriestandaarden en of deze geschikt is voor advocatuurlijke dienstverlening.

Het is echter risicovol om uitsluitend op de aan- of afwezigheid van een certificaat of auditrapport te varen. Dergelijke certificaten en rapporten vormen geen vervanging van deugdelijk eigen onderzoek en doorlopend toezicht op de dienstverlening. De interpretatie van dergelijke stukken kan bovendien lastig zijn en de waarde ervan staat of valt in beginsel met drie elementen: (i) de reikwijdte van het onderzoek (welke diensten, systemen of processen zijn daadwerkelijk getoetst); (ii) het moment waarop het onderzoek is uitgevoerd (een momentopname of rapport van een jaar geleden zegt weinig over de actuele situatie); en (iii) de reikwijdte van de onderzochte entiteiten (of het rapport ziet op de leverancier zelf, op een specifiek onderdeel of vestiging daarvan, of ook op de onderliggende modelleverancier en eventuele onderaannemers).
Een aantal praktische tips voor advocaten:

- Vraag niet alleen of een certificaat of rapport bestaat, maar ook welke onderdelen van de dienst daadwerkelijk zijn getoetst, en vraag het volledige rapport op in plaats van genoegen te nemen met alleen het certificaat of samenvatting.
- Controleer de geldigheidsdatum en de onderzoeksperiode: een verouderd rapport zegt weinig over de actuele stand van zaken.
- Ga na of het rapport ziet op de specifieke dienst die wordt afgenomen, of op de leverancier in bredere zin, en vraag door als dit niet duidelijk is.
- Vraag expliciet of de certificering of het auditrapport ook de onderliggende modelleverancier(s) en een eventuele tussenliggende lagen dekt.
- Schakel bij twijfel over de interpretatie van een certificaat of auditrapport gespecialiseerde ondersteuning in, net als bij overige contractbeoordeling.
- Gebruik een certificaat of rapport als een relevant gegeven bij de oriëntatie- en demofase, maar laat dit nooit de plaats innemen van een eigen, actieve verificatie van kritieke waarborgen tijdens de pilot of van periodieke controle na ingebruikname.

C.3 Beoordeel zorgvuldig of een Data Protection Impact Assessment nodig is

1 Aanbeveling

De advocaat voert een Data Protection Impact Assessment (DPIA) uit wanneer het gebruik van een AI-tool waarschijnlijk een hoog risico oplevert voor de rechten en vrijheden van betrokkenen in de zin van artikel 35 AVG.

2 Toelichting

Advocaten verwerken in veel gevallen persoonsgegevens en artikel 35 AVG vereist een DPIA bij verwerkingen die waarschijnlijk een hoog risico opleveren.

C.4 Wees alert op fraude bij cliëntidentificatie

1 Aanbeveling

Verifieer de identiteit van cliënten via een onafhankelijk kanaal en vertrouw niet uitsluitend op videocontact of digitale documenten als bewijs van identiteit. Wees voortdurend alert op de toenemende mogelijkheden voor kwaadwillenden om zich voor te doen als een cliënt.

2 Toelichting

AI maakt het mogelijk realistische vervalsingen te genereren van beelden en documentatie van personen en organisaties, wat fraude vergemakkelijkt. Dit raakt onze vertrouwelijkheid en de integriteit. Deze aanbeveling vereist niet dat elke identiteitscontrole via een fysieke ontmoeting verloopt, maar wel dat niet uitsluitend op één, mogelijk manipuleerbaar kanaal wordt vertrouwd. Het risico weegt zwaarder bij volledig op afstand tot stand gekomen cliëntrelaties zonder eerder persoonlijk contact en bij transacties met een verhoogd witwas- of fraude-risico. Wees extra alert bij ongebruikelijke haast, afwijkende betaalinstructies, of onduidelijkheden in de verstrekte documentatie, die kunnen wijzen op AI-geassisteerde fraude.

D Transparantie en verantwoording

D.1 Informeer cliënten over het AI-gebruik en vraag toestemming waar nodig

1 Aanbeveling

Informeer cliënten standaard over het kantoorbrede AI-beleid. Vraag vooraf toestemming in concrete gevallen zodra redelijkerwijs te verwachten is dat een geïnformeerde cliënt bezwaar zou maken tegen het gebruik van AI in diens dossier. Voorafgaande toestemming vragen wordt aanbevolen: (a) wanneer de advocaat niet het eindresultaat handmatig beoordeelt; (b) wanneer het beroepsgeheim daarmee niet gewaarborgd kan worden; (c) wanneer de cliënt eerder blijk heeft gegeven van bezwaren tegen AI-gebruik; (d) bij de verwerking van bijzondere persoonsgegevens; (e) bij situaties waarbij de advocaat kennis heeft dat de cliënt met derden een geheimhoudingsovereenkomst heeft gesloten die verwerking met behulp van AI (of het delen van informatie met derden) verbiedt; en (f) wanneer publieke modellen getraind worden op basis van de zaakgegevens; en (g) wanneer het kantoorbrede AI-beleid anderszins toestemming voorschrijft.

2 Toelichting

Neem een standaardparagraaf over AI-gebruik op in de opdrachtbevestiging of de algemene voorwaarden en informeer cliënten over derden die betrokken zijn bij de AI-gegevensverwerking en de waarborgen die het kantoor standaard treft (e.g. verbieden van training publieke modellen met cliëntgegevens, passende technische en organisatorische maatregelen, etc.). Toets bij intake ook of de cliënt eigen richtlijnen hanteert voor toeleveranciers met een AI-verbod of toestemmingsvereiste, en leg een eventuele beperking vast in het zaaksbeheersysteem en maak dit op passende wijze duidelijk aan behandelend advocaten. Documenteer concrete situaties waarin om toestemming gevraagd wordt. Zo moet het de advocaat bijvoorbeeld vrijstaan om in het kader van due diligence-werkzaamheden met de cliënt af te spreken dat AI gebruikt wordt om grote hoeveelheden overeenkomsten te controleren op termination for change of control-clausules. Daarbij kan worden afgesproken dat slechts 20% van de resultaten handmatig gecontroleerd wordt, en dat bij ten minste 90% validatie van die steekproef de overige 80% niet handmatig gecontroleerd wordt. Als iedere bevinding handmatig gecontroleerd zou moeten worden, zou dat de kostenefficiëntie in de weg staan. Maar omdat afgeweken wordt van de handmatige verificatieplicht van eindresultaten, wordt aanbevolen om de cliënt te informeren over de mogelijke risico's en te vragen om toestemming. Deze toestemming moet in het dossier op passende wijze gedocumenteerd worden door de advocaat. Deze norm vloeit voort uit de bescherming van de advocatuurlijke kernwaarden integriteit en vertrouwelijkheid en vormt geen vervanging of samenvatting van de informatievereisten of de vereisten tot het bepalen van de juiste grondslag die reeds gelden op grond van toepasselijke privacywetgeving. Essentieel is dat deze norm haalbaar is met inachtneming van de voorspelling dat onze beroepsgroep vaker AI wel dan niet zal gebruiken.

D.2 Wees intern transparant over het gebruik van AI

1 Aanbeveling

De advocaat is binnen het kantoor open over wanneer en hoe AI is gebruikt bij een advies, document of processtuk. Moedig een verantwoorde cultuur van experimenteren, het delen van ervaringen, en kritische evaluatie aan.

2 Toelichting

Dit raakt integriteit en bevordert kwaliteitscontrole en collegiale toetsing. Interne transparantie, gekoppeld aan opleidingsmomenten, bevordert de kwaliteit van de output. Het gebruik van AI mag daarom geen taboe zijn en moet nooit op zichzelf stigmatiserend werken. Die transparantie is juist noodzakelijk om de output van AI goed te kunnen controleren. Waar het ontbreken van interne transparantie herhaaldelijk voorkomt of tot kwaliteitsproblemen leidt, kan dit, naast een leermoment, in uiterste gevallen ook aanleiding geven tot interne sancties conform het reguliere personeelsbeleid van het kantoor.

D.3 Meld AI-gebruik aan de rechter waar dit redelijkerwijs aangewezen is

1 Aanbeveling

De advocaat meldt het gebruik van AI bij het opstellen van een processtuk aan de rechter wanneer dit gebruik van dien aard is dat wetenschap van het gebruik daarvan redelijkerwijs relevant is voor de beoordeling van het stuk.

2 Toelichting

Dit raakt onder meer de integriteit. Deze norm vereist geen algemene meldplicht voor ieder gebruik van AI-ondersteuning. De verwachting van de projectgroep is juist dat AI-gebruik de regel wordt en niet de uitzondering en bovendien oordelen procespartijen en rechters in beginsel op de merites van een stuk, niet op de wijze van totstandkoming. De informatieplicht moet beperkt zijn tot situaties waarin de schaal of aard van het AI-gebruik een reëel risico met zich meebrengt op benadeling van een procespartij of misleiding van de rechter. Denk hier, bijvoorbeeld, aan de inzet van AI met als doel het extra omvangrijk maken van een processtuk of aan geautomatiseerde strategieën om in korte tijd grote hoeveelheden procedures aanhangig te maken op basis van hoofdzakelijk door AI-gegenereerde standaardstukken of het indienen van AI-gegenereerde foto's als productie in een letselschadezaak.

D.4 Declareer alleen daadwerkelijke tijd wanneer op basis van uurtarieven wordt gefactureerd

1 Aanbeveling

De advocaat declareert op basis van daadwerkelijk verrichte werkzaamheden, ook wanneer AI een deel van het werk in aanzienlijk kortere tijd heeft voortgebracht.

2 Toelichting

Wanneer een advies dat met AI in enkele uren is gegenereerd, wordt gedeclareerd alsof er sprake was van dagen handmatig werk, raakt dit de integriteit van de dienstverlening; cliënten dienen niet te betalen voor uren die niet zijn gemaakt. Dit raakt de integriteit van onze beroepsgroep. Deze aanbeveling strekt allerminst tot het verbieden van andere verdienmodellen dan uurtarief. Sterker nog, deze aanbeveling moet worden gezien als een uitnodiging aan de beroepsgroep om de waarde van, en benodigde investering in, AI – maar eigenlijk ook IT in algemene zin - kritisch te evalueren om te bepalen hoe dit doorwerkt in commerciële modellen en hoe commerciële modellen beter gericht kunnen worden op de waarde die zij opleveren voor cliënten.

Essentieel is dat de advocaat met de cliënt bespreekt, waar relevant, hoe AI-ondersteuning van invloed is op de kosten en dat alleen de tijd gedeclareerd wordt die daadwerkelijk is besteed aan het opstellen, beoordelen en verfijnen van AI-ondersteunde output. Dit laat onverlet dat advocaten vrij zijn om vaste prijsafspraken te maken waarin de kosten van AI zijn verdisconteerd.

E Partijdigheid en maatschappelijke impact

E.1 AI mag nooit leidend zijn in de partijdige belangenbehartiging

1 Aanbeveling

De advocaat blijft verantwoordelijk voor de onafhankelijkheid van diens oordeel en een partijdige belangenbehartiging. AI mag hierin nooit doorslaggevend zijn.

2 Toelichting

Dit raakt onder meer de kernwaarde partijdigheid. AI-systemen genereren op basis van waarschijnlijkheden en gemiddelden, terwijl partijdige belangenbehartiging juist creativiteit, maatwerk en soms een afwijkende koers vraagt. Deze norm sluit het gebruik van AI ter inspiratie of ondersteuning niet uit, zolang de advocaat de uiteindelijke inhoudelijke koers zelf bepaalt. Wees alert op de neiging van AI om tot een ‘standaardoutput’ te komen; het gebruik van doordachte instructies en prompting frameworks kan dit risico beperken.

E.2 Wees alert op algoritmische bias

1 Aanbeveling

Beoordeel kritisch of AI de individuele omstandigheden van de cliënt en de zaak voldoende meeweegt.

2 Toelichting

AI-systemen zijn getraind op historische data en kunnen bestaande ongelijkheden reproduceren of versterken. Dit raakt primair partijdigheid, omdat vooringenomen output het risico op benadeling van (groepen) cliënten of wederpartijen met zich meebrengt. Deze aanbeveling roept op tot kritische toets vóór gebruik (e.g. bij beoordeling van het AI-systeem) en daarna tot doorlopende alertheid bij de beoordeling van de output.

Wees met name alert wanneer AI-systemen grondig gewijzigd worden (i.e. zgn. “new releases” of “major upgrades”). Dit weegt zwaarder naarmate de uitkomst van de AI-toepassing directer van invloed is op de individuele positie van de cliënt (bijvoorbeeld bij advisering over proceskansen) en minder zwaar bij algemene informatie die geen betrekking heeft op zaaksgegevens. Beoordeel bij het gebruik van AI voor advies over proceskansen of individuele beoordelingen steeds of de output aansluit bij de specifieke, individuele omstandigheden van het geval. Wees in het bijzonder alert bij grensoverschrijdend of meertalig gebruik: generieke AI-tools beantwoorden een in het Nederlands gestelde vraag doorgaans op basis van Nederlands recht, ook wanneer een andere jurisdictie is bedoeld, en blijven dat geneigd te doen zelfs nadat de juiste jurisdictie expliciet is aangegeven. Verifieer daarom bij grensoverschrijdende kwesties uitdrukkelijk of de output daadwerkelijk op het toepasselijke rechtsstelsel is gebaseerd. Wees bedacht op onverwachte of opvallend eenzijdige antwoorden als mogelijke indicator van bias.

E.3 Houd er rekening mee dat AI kan leiden tot ongelijkheid in de toegang tot het recht.

1 Aanbeveling

Wees alert op situaties waarin de wederpartij beschikt over kwalitatief betere AI-middelen, en op de mogelijkheid dat AI de belangenbehartiging van de eigen cliënt juist kan versterken ten opzichte van een grotere wederpartij. Kleinere kantoren wordt aangeraden om de mogelijkheden te verkennen om meer te gaan samenwerken op het gebied van inkoop van AI-software.

2 Toelichting

AI kan verschillen tussen grote en kleine kantoren zowel vergroten als verkleinen. Geavanceerde AI-tools zijn vaak kostbaar en niet even toegankelijk voor grote en kleine kantoren. Dit raakt partijdigheid in de zin van gelijkwaardige procesvoering. Tegelijk kan een kleiner kantoor dat slim gebruik maakt van minder geavanceerde AI beter partij bieden aan een groter kantoor die niet efficiënt gebruikmaakt van geavanceerde software. Kleinere kantoren kunnen overwegen hun koopkracht te bundelen en de kosten van IT-inkoop en beheer met elkaar te delen.

E.4 Waak voor proceduremisbruik via AI

1 Aanbeveling

Beoordeel of het gebruik van AI in een concreet dossier doelmatig is, in overeenstemming is met de beginselen van een eerlijk proces en gebruik AI niet bewust om procespartijen of rechters te overvallen met informatie die niet relevant is voor de rechtsdiscussie.

2 Toelichting

AI heeft de neiging om breedsprakig te zijn en veelvuldig hetzelfde te herhalen bij het genereren van teksten. Het moeten doornemen van enorme hoeveelheden irrelevante informatie kan de toegang tot het recht vertragen en kostbaarder maken. Dit raakt partijdigheid in combinatie met integriteit. Deze aanbeveling strekt niet tot het verbieden van efficiënte inzet van AI bij omvangrijke dossiers, maar roept wel op terughoudendheid bij het opvoeren van omvang of complexiteit met als doel procespartijen of rechtspraak te overbelasten. Dit weegt zwaarder naarmate de omvang of frequentie van AI-gegenereerde stukken sterk afwijkt van wat voor het type zaak gebruikelijk is, en houdt direct verband met de eerdergenoemde aanbeveling tot transparantie.

E.5 Benut AI voor bredere toegang tot het recht

1 Aanbeveling

Breng in kaart hoe AI-gebruik binnen de eigen praktijk kan bijdragen aan toegankelijker en betaalbaarder juridische dienstverlening voor (mogelijke) cliënten die momenteel onvoldoende toegang hebben tot dergelijke juridische bijstand.

2 Toelichting

Dit raakt partijdigheid in de bredere, maatschappelijke betekenis: AI biedt niet alleen risico's maar ook kansen om de toegang tot het recht te vergroten. Overweeg hoe AI-ondersteunde efficiëntie kan worden vertaald in lagere kosten of grotere beschikbaarheid van rechtsbijstand, bijvoorbeeld binnen de sociale advocatuur of pro-bono praktijk. Denk bijvoorbeeld aan het aanbieden van een chatbot die mensen die het Nederlands niet machtig zijn helpt bij dossierintake of een chatbot die cliënten helpt om sneller de meest geschikte advocaat binnen een kantoor te vinden.

E.6 Wees alert op belangenconflicten bij gedeelde AI-systemen

1 Aanbeveling

De advocaat beoordeelt of het gebruik van een AI-systeem met toegang tot informatie van meerdere cliënten van het kantoor (e.g. via gedeelde contextvensters, geheugenfuncties, of finetuning op kantoorbrede data) een risico op ongewenste informatie-uitwisseling tussen cliënten of belangenverstrengeling met zich meebrengt, en treft passende technische of organisatorische waarborgen wanneer dat risico zich kan voordoen.

2 Toelichting

AI-systemen die door het kantoor breed worden ingezet, kunnen worden getraind op, of tijdens gebruik toegang hebben tot, gegevens van meerdere cliënten tegelijk. Dit kan ertoe leiden dat informatie van de ene cliënt ongemerkt doorwerkt in de advisering aan een andere cliënt, bijvoorbeeld via een gedeeld geheugen van een AI-tool, een kantoorbrede kennisbank die als context wordt meegegeven, of een model dat specifiek is gefinetuned op eerdere dossiers van het kantoor. Dit risico kan zich ook voordoen wanneer het AI-systeem uitsluitend gegevens van het eigen kantoor gebruikt als twee partijen allebei cliënt zijn geweest bij het kantoor en op enig moment tegenover elkaar komen te staan. Dit raakt de vertrouwelijkheid en de partijdigheid. Overweeg daarom bij de inrichting en inkoop van AI-systemen met een gedeeld geheugen of kantoorbrede kennisbank of, en zo ja hoe, technische scheiding tussen dossiers van verschillende cliënten geborgd is, en betrek dit onderwerp ook in de reguliere conflict-of-interestcontrole van het kantoor.

Annex 1 – Toelichting technische begrippen

Deze begrippenlijst licht technische termen toe die in de aanbevelingen en de bijbehorende toelichtingen worden gebruikt. De begrippen zijn alfabetisch gerangschikt. Waar de hoofdtekst een Engelse benaming gebruikt, is die bij de Nederlandse uitleg vermeld. De toelichtingen zijn bedoeld als leeshulp, niet als wettelijke definities of als vervanging van concrete afspraken met een leverancier. Er is bewust gekozen om bepaalde nuance achterwege te laten wanneer dit voor de niet-technische lezer voor meer vragen of onduidelijkheid zou zorgen. De toelichting op de begrippen is om die reden vooral gericht op de context van de aanbevelingen.

- **Adapter-bestand.** Een aanvullend bestand waarmee het gedrag van een bestaand AI-model voor een bepaalde taak of organisatie wordt aangepast, zonder het oorspronkelijke model opnieuw te trainen. Zo'n bestand kan informatie uit de gebruikte trainingsgegevens bevatten en verdient daarom ook aandacht bij beveiliging en verwijdering.

- **Agentic AI (AI-agent).** AI die binnen een gegeven opdracht zelf meerdere stappen kan plannen en uitvoeren, en daarvoor – waar nodig – ook zelfstandig bij andere programma's of diensten kan inloggen en functionaliteit kan aanroepen. Een agent kan bijvoorbeeld documenten ophalen, bewerken en een e-mail voorbereiden of versturen, afhankelijk van de verleende bevoegdheden.

- **AI (artificiële of kunstmatige intelligentie).** Een verzamelnaam voor computertechnieken waarmee systemen bijvoorbeeld patronen herkennen, voorspellingen doen of teksten en andere inhoud maken. AI is niet hetzelfde als menselijk begrip en levert niet vanzelf juiste of betrouwbare resultaten.

- **AI-model en AI-systeem (of AI-tool).** Een AI-model is het getrainde rekenmodel dat bijvoorbeeld tekst genereert of informatie indeelt. Een AI-systeem of AI-tool is het programma waarmee de gebruiker werkt; die tool kan één of meer modellen combineren met zoekfuncties, documentopslag en andere diensten. Om het onderscheid te onthouden, kan gedacht worden aan een auto waarbij de motor het AI-model (met optimalisatie) voorstelt en de rest van de auto het AI-systeem of de AI-tool.

- **Anonimisering.** Het zodanig bewerken van gegevens dat personen niet meer redelijkerwijs identificeerbaar zijn. Alleen namen weghalen is vaak onvoldoende: ook de combinatie van feiten in een dossier kan een persoon herkenbaar maken.

- **API (Application Programming Interface).** Een technische koppeling waarmee programma's gegevens uitwisselen of elkaars functies gebruiken. Een AI-tool kan via een API bijvoorbeeld een taalmodel of een externe databank raadplegen; daarbij kunnen gegevens naar een andere dienstverlener gaan.

- **Authenticatie, autorisatie en IAM.** Authenticatie controleert wie iemand is, bijvoorbeeld bij het inloggen. Autorisatie bepaalt wat die persoon mag zien of doen. IAM (Identity and Access Management) is het geheel van voorzieningen waarmee een organisatie identiteiten en toegangsrechten beheert.

- **Back-up.** Een reservekopie van gegevens die herstel na verlies of een storing mogelijk maakt. Gegevens verwijderen uit een toepassing betekent niet automatisch dat zij ook direct uit alle back-ups zijn verdwenen.

- **Benchmark en testdataset.** Een benchmark is een vergelijking van prestaties aan de hand van vaste taken en beoordelingscriteria. Een testdataset is de verzameling voorbeelden waarmee die prestaties worden getoetst, bijvoorbeeld contracten met vooraf gecontroleerde antwoorden; goede testresultaten garanderen niet dat ieder nieuw antwoord juist is.

- **Bias (algoritmische bias of vertekening).** Een systematische scheefheid in de werking of uitkomsten van een AI-systeem, bijvoorbeeld doordat bepaalde groepen of rechtsstelsels in de gebruikte gegevens over- of ondervertegenwoordigd zijn. 'Algoritmisch' verwijst naar de rekenregels en verwerkingsstappen van het systeem. Hierdoor kan een antwoord eenzijdig zijn of bepaalde personen benadelen.

- **Caching.** Het tijdelijk bewaren van gegevens of eerdere berekeningen om een volgende verwerking sneller of goedkoper te maken. Ook tijdelijk bewaarde prompts of documentfragmenten kunnen vertrouwelijke informatie bevatten; bewaarlocatie en bewaartermijn zijn daarom relevant.

- **Context en contextvenster.** Context is de informatie die een AI-model bij een concrete opdracht krijgt, zoals instructies, gespreksgeschiedenis en documentfragmenten. Het contextvenster is de maximale hoeveelheid informatie die het model bij één verwerking kan meenemen. Een groot document uploaden betekent daarom niet vanzelf dat het model ieder onderdeel tegelijk gebruikt.

- **Data-extractie en afleiding van gegevens (data extraction/inference).** Het achterhalen van informatie uit een AI-systeem, bijvoorbeeld door vertrouwelijke tekst te laten reproduceren of gevoelige kenmerken af te leiden uit antwoorden.

- **Dataminimalisatie.** Alleen de gegevens gebruiken die voor het concrete doel nodig zijn. Bij een vraag over één contractsbepaling kan het bijvoorbeeld voldoende zijn die bepaling met de noodzakelijke achtergrond te delen, in plaats van het volledige dossier.

- **Dataportabiliteit en machineleesbaar formaat (machine-readable formaat).** Dataportabiliteit is de mogelijkheid om gegevens mee te nemen naar een andere toepassing of leverancier. Een machineleesbaar formaat, zoals CSV, JSON of XML, maakt geautomatiseerde verwerking mogelijk; of de gegevens elders daadwerkelijk bruikbaar zijn, hangt ook af van de structuur en beschikbare export- en importmogelijkheden.

- **Data- en contextvergiftiging (data poisoning).** Het manipuleren van trainingsgegevens of bronnen die een AI-systeem gebruikt, zodat de uitkomsten worden beïnvloed. Denk aan een gemanipuleerd document in een kennisbank waardoor het systeem onjuiste informatie als uitgangspunt neemt.

- **Embeddings.** Reeksen getallen die kenmerken van bijvoorbeeld een tekst weergeven, zodat een systeem teksten op inhoudelijke overeenkomst kan doorzoeken. Zij zijn niet hetzelfde als de oorspronkelijke tekst, maar zijn ook niet vanzelf anoniem of vrij van vertrouwelijke informatie.

- **Encryptie (versleuteling) en cryptografische sleutel.** Encryptie maakt gegevens onleesbaar voor wie niet over de juiste sleutel beschikt. Versleuteling 'in rust' beschermt opgeslagen gegevens; versleuteling 'in transport' beschermt gegevens tijdens verzending. Van belang is ook wie de sleutel beheert en wie de gegevens tijdens gebruik kan inzien.

- **Fine-tuning.** Het verder trainen van een bestaand AI-model met aanvullende voorbeelden om het voor een bepaalde taak of werkwijze aan te passen. Dit wijzigt het model en is iets anders dan documenten uitsluitend bij een afzonderlijke opdracht als context meegeven.

- **Firewall, netwerkisolatie en DMZ.** Een firewall regelt welk netwerkverkeer wordt toegelaten. Netwerkisolatie scheidt systemen van elkaar om toegang en de verspreiding van aanvallen te beperken. Een DMZ is een afzonderlijk netwerkdeel voor systemen die van buiten bereikbaar moeten zijn, afgeschermd van het interne netwerk.

- **Gegevensisolatie.** Het technisch gescheiden houden van gegevens en toegang, bijvoorbeeld tussen kantoren, cliënten of dossiers. Alleen verschillende mappen of schermen tonen is niet voldoende: de scheiding moet ook in de achterliggende systemen en toegangsrechten werken.
- **Generatieve AI.** AI die op basis van een opdracht nieuwe inhoud maakt, zoals tekst, beelden, audio of video. De inhoud kan overtuigend zijn zonder feitelijk juist te zijn of op een bestaande bron te berusten.
- **Hallucinatie.** Een door AI gegenereerde onjuiste of verzonnen mededeling die als juist wordt gepresenteerd, zoals een niet-bestaande uitspraak, een onjuist citaat of een verzonnen dossierfeit. Een stellig antwoord of een toegevoegde bronverwijzing sluit een hallucinatie niet uit.
- **Inferentie.** Het gebruiken van een al getraind AI-model om een concrete invoer te verwerken en een resultaat te produceren. De plaats waar dit gebeurt kan verschillen van de plaats waar de documenten worden opgeslagen.
- **Input en output.** Input is wat aan een systeem wordt aangeboden, zoals een vraag, document, afbeelding of geluidsbestand. Output is wat het systeem teruggeeft, bijvoorbeeld een samenvatting, antwoord of concepttekst.
- **Jailbreaking.** Een poging om een AI-systeem met speciaal geformuleerde opdrachten zijn ingebouwde beperkingen of veiligheidsregels te laten omzeilen. Dit kan bijvoorbeeld gericht zijn op het verkrijgen van informatie die het systeem niet hoort prijs te geven.
- **Least privilege.** Het beginsel dat een gebruiker of systeem alleen de toegang en bevoegdheden krijgt die voor de taak noodzakelijk zijn. Een AI-agent die een concept moet maken, hoeft bijvoorbeeld niet ook zelfstandig e-mails te kunnen verzenden.
- **Logging en monitoring.** Logging is het vastleggen van gebeurtenissen, zoals inloggen, het raadplegen van documenten of acties van een AI-agent. Monitoring is het volgen en beoordelen daarvan om problemen te signaleren. Ook logbestanden kunnen vertrouwelijke gegevens bevatten.
- **Metadata.** Gegevens over andere gegevens of over het gebruik van een systeem, zoals een bestandsnaam, auteur, tijdstip of IP-adres. Ook zonder de inhoud van een document kunnen metadata gevoelige informatie prijsgeven, bijvoorbeeld over een cliëntrelatie.
- **Middleware.** Software die als tussenlaag verschillende systemen met elkaar verbindt, bijvoorbeeld een advocatentoepassing met een externe modelleverancier. Ook deze tussenlaag kan gegevens verwerken of bewaren en is daarom onderdeel van de te beoordelen gegevensketen.
- **Modelgewichten.** De getallen in een AI-model die tijdens training worden aangepast en samen bepalen hoe het model reageert. Het zijn geen gewoon leesbare dossierbestanden, maar zij kunnen wel informatie uit trainingsgegevens bevatten.
- **Multi-factor-authenticatie (MFA).** Inloggen met ten minste twee verschillende soorten bewijs van identiteit, bijvoorbeeld een wachtwoord en een beveiligingssleutel. Phishingresistente MFA is zo ingericht dat een nagemaakte inlogpagina de inlogmiddelen niet eenvoudig kan onderscheppen en hergebruiken.

- **Open source.** Bij opensourcesoftware is de broncode beschikbaar onder voorwaarden die gebruik, aanpassing en verspreiding toestaan. Controleer bij een AI-toepassing welke onderdelen daadwerkelijk onder die voorwaarden beschikbaar zijn. Open source betekent niet automatisch dat een toepassing veilig is of lokaal draait.

- **OTAP.** Afkorting van ontwikkeling, test, acceptatie en productie: gescheiden omgevingen om software te bouwen, te controleren en uiteindelijk te gebruiken. Deze scheiding beperkt het risico dat een ongeteste wijziging de dagelijkse dienstverlening raakt.

- **Patching en systeemhardening.** Patching is het installeren van updates om onder meer beveiligingsfouten te herstellen. Systeemhardening is het veiliger instellen van een systeem, bijvoorbeeld door onnodige functies en verbindingen uit te schakelen.

- **Penetratietest (pentest), kwetsbaarheidsscan en security audit.** Een kwetsbaarheidsscan zoekt naar mogelijke beveiligingszwaktes; een pentest onderzoekt door toegestane aanvalspogingen of zwaktes daadwerkelijk kunnen worden misbruikt. Een security audit beoordeelt beveiligingsmaatregelen en processen. De betekenis van elk onderzoek hangt af van wat, wanneer en hoe is getest.

- **Prompt, prompting en promptbibliotheek (prompt library; meervoud: prompt libraries).** Een prompt is een vraag of instructie aan een AI-systeem. Prompting is het formuleren en verfijnen daarvan; een promptbibliotheek is een verzameling herbruikbare instructies. Een goede prompt kan de bruikbaarheid vergroten, maar garandeert geen juist antwoord.

- **Prompt-injectie (prompt injection).** Een aanval waarbij instructies in bijvoorbeeld een document, e-mail of website het AI-systeem proberen af te leiden van de eigenlijke opdracht. Een document dat de agent moet samenvatten kan bijvoorbeeld verborgen instructies bevatten om vertrouwelijke informatie door te sturen.

- **Pseudonimisering.** Het vervangen van rechtstreeks identificerende gegevens door bijvoorbeeld codes, waarbij identificatie met aanvullende informatie nog mogelijk is. Dit verschilt van anonimisering; ook de resterende dossierinhoud kan een persoon herkenbaar maken.

- **RAM en vRAM (werkgeheugen).** Geheugen waarin een computer of grafische processor gegevens tijdens de verwerking tijdelijk beschikbaar houdt. Dit verschilt van opslag op een schijf; tijdelijke verwerking sluit op zichzelf niet uit dat gegevens daarnaast in logs, caches of andere opslag terechtkomen.

- **Retrieval en RAG (Retrieval-Augmented Generation).** Retrieval is het ophalen van relevante informatie uit bronnen. Bij RAG krijgt een generatief AI-model zulke opgehaalde informatie mee om een antwoord te maken, bijvoorbeeld passages uit een dossier of juridische databank. Dit kan de aansluiting op bronnen verbeteren, maar voorkomt geen zoekfouten, onjuiste interpretaties of hallucinaties.

- **Sycophancy.** De neiging van een AI-systeem om aan te sluiten bij de kennelijke mening of verwachting van de gebruiker, ook als die onjuist is. Een advocaat die vraagt waarom diens standpunt juist is, kan daardoor bevestiging krijgen in plaats van een kritische toets.

- **Systeeminstructie.** Een vooraf ingestelde instructie die het gedrag van een AI-systeem stuurt, bijvoorbeeld over de rol, werkwijze of te gebruiken bronnen. Een systeeminstructie is geen harde technische beveiligingsgrens en garandeert niet dat het systeem zich er steeds aan houdt.

- **Taalmodel en LLM (Large Language Model).** Een AI-model dat op grote hoeveelheden gegevens is getraind om taalpatronen te verwerken en tekst te genereren. Een LLM maakt tekst op basis van aangeleerde patronen en meegegeven context; het is niet vanzelf een actuele juridische databank.

- **TLS en HTTPS.** Technieken voor het beveiligen van gegevens tijdens transport via een netwerk; HTTPS gebruikt TLS om de verbinding met een website te beschermen. Een beveiligde verbinding zegt niet of de ontvanger de gegevens daarna zorgvuldig bewaart of gebruikt.

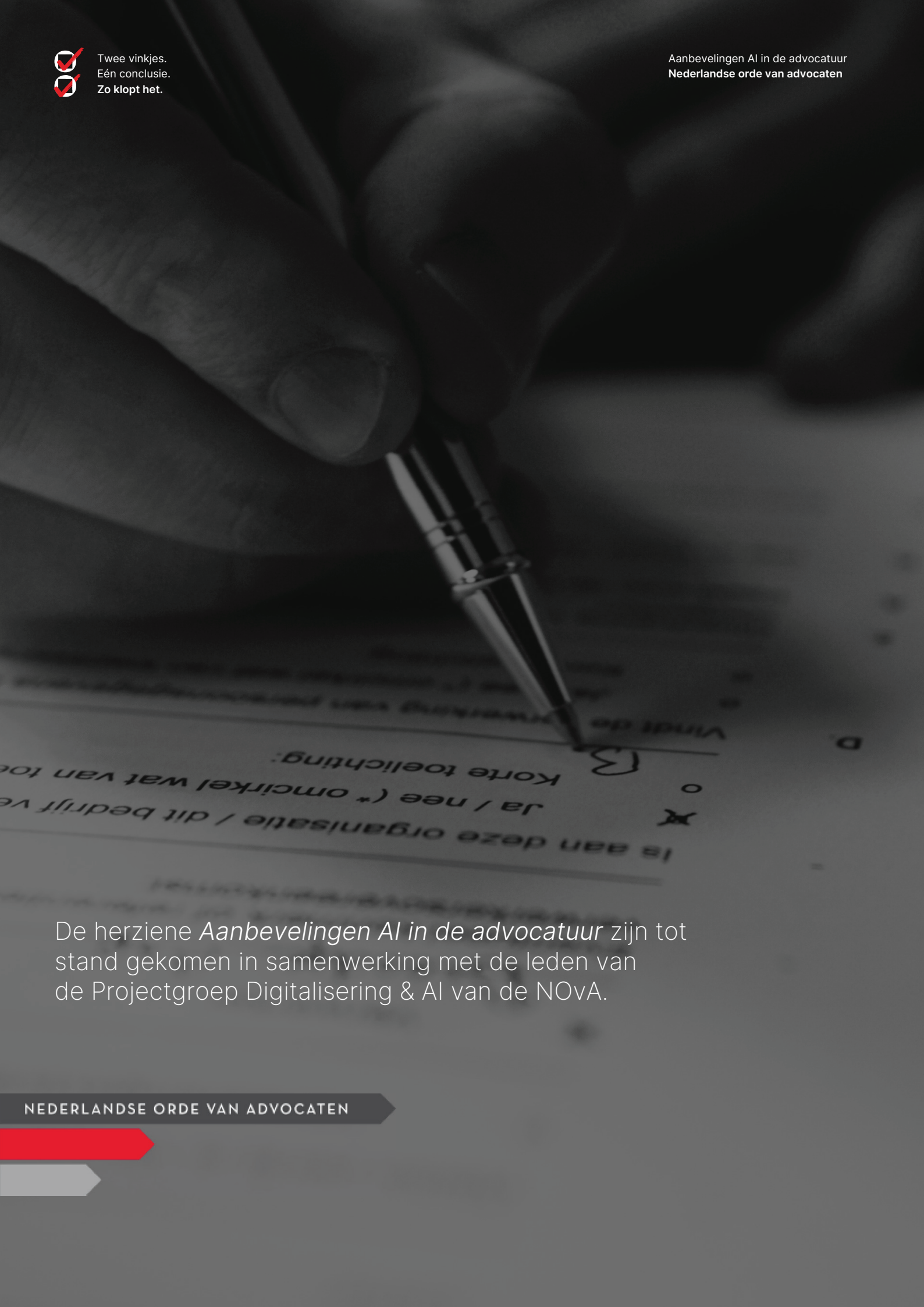
- **Training, pre-training en RLHF.** Training is het aanpassen van een model aan de hand van voorbeelden. Pre-training is de brede basistraining; RLHF (Reinforcement Learning from Human Feedback) gebruikt menselijke beoordelingen om het gedrag van een model bij te sturen. Geen training op cliëntgegevens betekent niet automatisch dat die gegevens ook niet worden opgeslagen.

- **Uptime en downtime.** Uptime is de tijd waarin een dienst beschikbaar is; downtime is de tijd waarin deze niet beschikbaar is. Een beschikbaarheidspercentage is pas goed te beoordelen als duidelijk is over welke periode wordt gemeten en welke storingen of onderhoudsperioden meetellen.

- **Vendor lock-in.** Afhankelijkheid van een leverancier waardoor overstappen moeilijk, duur of tijdrovend wordt, bijvoorbeeld omdat gegevens niet bruikbaar kunnen worden geëxporteerd of werkprocessen nauw met de dienst verweven zijn.

- **Whitelist (lijst van goedgekeurde toepassingen).** Een overzicht van toegestane programma's of diensten. Bij AI is ook van belang voor welke taken en gegevens een toepassing is goedgekeurd; opname op de lijst betekent niet dat ieder gebruik is toegestaan. Het alternatief is een "blacklist" waarbij wordt bepaald welke programma's of diensten niet zijn toegestaan.

- **Zero Data Retention (ZDR).** Voor het doel van deze aanbevelingen: de technische en contractuele toezegging dat de leverancier input en output alleen tijdelijk voor de gevraagde verwerking gebruikt en daarna wist, zonder blijvende opslag, logging van de inhoud, hergebruik voor training of caching over sessies heen. De precieze reikwijdte, eventuele uitzonderingen en toepasselijkheid op onderliggende leveranciers moeten uitdrukkelijk worden gecontroleerd; het etiket ZDR alleen is onvoldoende.



De herziene *Aanbevelingen AI in de advocatuur* zijn tot stand gekomen in samenwerking met de leden van de Projectgroep Digitalisering & AI van de NOvA.

NEDERLANDSE ORDE VAN ADVOCATEN