

AI in de advocatuur: basisbegrippen en veelgestelde vragen

Basisbegrippen

AI (Kunstmatige Intelligentie)

Technologie die taken uitvoert waarvoor normaal menselijke intelligentie nodig is.

Bias

Vooringenomenheid of systematische fout in de AI, vaak veroorzaakt door eenzijdige of onvolledige trainingsdata. Dit kan leiden tot oneerlijke of onjuiste uitkomsten.

Bronvermelding

Verwijzing naar de herkomst van informatie; essentieel voor het controleren van AI-output.

Exit-clausule

Afspraken over wat er gebeurt als je wilt stoppen met een AI-dienst, bijvoorbeeld over data-eigendom, overdraagbaarheid en beëindiging van toegang.

Generatieve AI (GenAI)

AI-systemen die zelfstandig nieuwe content genereren, zoals tekst, beeld, audio of video. Zij doen dit op basis van grote hoeveelheden voorbeelden uit bestaande data.

Hallucinatie

Wanneer AI iets verzint dat niet waar is, zoals een niet-bestaande rechtszaak of fictieve jurisprudentie. Dit kan leiden tot misleidende of foutieve output.

Input / Output

Input is wat je invoert in een AI-systeem (de gegevens of vragen); output is het resultaat dat het systeem genereert (bijvoorbeeld een tekst of samenvatting).

LLM (Large Language Model)

Een type AI dat getraind is op grote hoeveelheden tekst en daardoor nieuwe teksten kan genereren die lijken op menselijke taal.

Prompt

De vraag of opdracht die je aan een AI-systeem geeft.

Prompt-engineering

Het slim formuleren van prompts om betere en betrouwbaardere AI-antwoorden te krijgen.

Sycophancy

De neiging van AI om het met de gebruiker eens te zijn, ook als dat niet terecht is. Dit kan leiden tot bevestiging van onjuiste aannames.

Vendor-lock-in

Situatie waarin je afhankelijk wordt van één leverancier van AI-diensten, waardoor overstappen naar een andere aanbieder lastig of kostbaar wordt.

Veelgestelde vragen

1. Algemeen

a. Is het gebruik van generatieve AI verplicht?

Nee. Het gebruik is niet verplicht, maar kan efficiënt zijn mits de juiste waarborgen worden toegepast.

b. Welke data worden verwerkt door een AI-tool?

Afhankelijk van de tool: ingevoerde tekst, metadata, locatiegegevens, gebruikersgedrag en mogelijk ook contextuele informatie zoals geopende tabs.

2. Veiligheid en privacy

a. Wat zijn datastromen?

De routes die gegevens afleggen, bijvoorbeeld van jouw computer naar de server van de AI-aanbieder. Belangrijk om te weten waar en door wie data wordt opgeslagen en verwerkt.

b. Wat is privacy-by-design?

Een aanpak waarbij privacy vanaf het begin wordt meegenomen in het ontwerp en gebruik van AI-tools. Deel alleen strikt noodzakelijke informatie en voorkom dat gevoelige data buiten je kantooromgeving terechtkomt.

c. Wat is een DPIA (Data Protection Impact Assessment)?

Een verplichte risicoanalyse bij het verwerken van persoonsgegevens in AI-systemen.

d. Wat is een datalek?

Wanneer vertrouwelijke informatie ongewenst openbaar wordt, bijvoorbeeld door een fout in een AI-tool.

3. Verantwoord gebruik

a. Mag ik vertrouwelijke cliëntstukken invoeren in een AI-tool zoals ChatGPT?

Alleen met stevige contractuele waarborgen of bij gebruik van on-premise oplossingen. Vermijd publieke AI-modellen zonder duidelijke garanties.

b. Ben ik verplicht AI-gebruik te melden aan mijn cliënt?

In geval van een datalek kan er een meldplicht zijn (zie Gedragsregel 16 en de AVG).

c. Wat is een AI-beleid?

Een set interne afspraken over het gebruik van AI binnen het kantoor, inclusief toegestane tools, verantwoordelijkheden, controlemechanismen en evaluatieprocedures.

d. Wat moet ik opnemen in een AI-beleid?

- Overzicht toegestane en verboden tools
- Rollen & verantwoordelijkheden
- AI-geletterdheidsbeleid
- Beoordelingscriteria
- Documentatie- en evaluatieplicht
- Verificatie- en escalatiemechanismen
- Uitvalscenario
- Periodieke audits

4. Kwaliteit en betrouwbaarheid

a. Waarom moet ik AI-output altijd controleren?

AI kan fouten maken of zelfs fictieve informatie genereren (hallucinaties). Controleer altijd citaten, jurisprudentie en feiten handmatig.

b. Hoe voorkom ik hallucinaties?

Volledig voorkomen van hallucinaties is niet mogelijk. AI-tools kunnen onjuiste, verzonden of misleidende informatie genereren, ook als ze overtuigend klinkt. Beperk risico's door:

- tools te gebruiken met bronvermelding
- bronnen handmatig te controleren
- output als concept te beschouwen
- zelf juridisch eindverantwoordelijk te blijven

c. AI voelt als een 'black box'; waar begin ik?

Begin klein, bijvoorbeeld met het samenvatten van publiek materiaal. Bouw ervaring op, volg trainingen en werk met een controleprotocol.

e. Is AI gewoon een zoekmachine?

Nee. Een LLM genereert tekst op basis van waarschijnlijkheid en kan onjuiste of verzonden informatie geven. Kritisch menselijk oordeel blijft essentieel.

5. Bias en partijdigheid

a. Wat is partijdigheid in AI-tools?

Wanneer een AI-tool bewust of onbewust de ene partij bevoordeelt boven de andere.

b. Hoe herken je bias en partijdigheid?

- Onverwachte of eenzijdige antwoorden: als de AI steeds dezelfde kant kiest of antwoorden geeft die opvallend aansluiten bij jouw vraagstelling, kan er sprake zijn van bias.
- Bevestiging van je eigen standpunt (sycophancy): AI-systemen, vooral grote taalmodellen, zijn geneigd om het met je eens te zijn, zelfs als je vraag niet klopt. Ze geven dan antwoorden die jouw mening bevestigen in plaats van kritisch te zijn.

- Onlogische of onrealistische uitkomsten: als de AI juridische argumenten of casussen verzint die niet bestaan, kan dat wijzen op een gebrek aan objectiviteit of op hallucinaties.
- Gebrek aan transparantie over hoe het antwoord tot stand kwam: als je niet kunt achterhalen hoe de AI tot een bepaald antwoord is gekomen, is het moeilijk om te beoordelen of er sprake is van bias.

6. AI-geletterdheid

a. Wat is AI-geletterdheid?

AI-geletterdheid verwijst naar het geheel van kennis, vaardigheden en attitudes die nodig zijn om AI op een verantwoorde, veilige, effectieve en kritische manier te begrijpen, gebruiken en beoordelen. Dit omvat kennis van AI-systemen, inzicht in regelgeving, risicobeoordeling, verantwoord gebruik en praktische vaardigheden.

In de context van de Europese AI-verordening (AI Act) betekent AI-geletterdheid dat individuen – en in het bijzonder professionals zoals advocaten – voldoende inzicht hebben in de werking, mogelijkheden, beperkingen en risico's van AI-systemen, zodat zij deze technologieën veilig en ethisch kunnen toepassen binnen hun beroepspraktijk.

b. Wat valt er onder AI-geletterdheid?

- *Kennis van AI-systemen*: begrijpen wat AI is, hoe AI-systemen (zoals generatieve AI) werken, en welke soorten AI er zijn.
- *Inzicht in regelgeving*: op de hoogte zijn van relevante wet- en regelgeving, waaronder de AI-Verordening en onder andere de AVG, en weten welke verplichtingen hieruit voortvloeien bij het gebruik van AI.
- *Beoordelen van risico's*: kunnen inschatten welke risico's AI-gebruik met zich meebrengt, zoals bias, hallucinaties, privacy- en dataproductie problemen, en het 'black box'-karakter van veel AI-systemen.
- *Verantwoord gebruik*: weten hoe AI op een veilige, transparante en ethisch verantwoorde manier kan worden ingezet, inclusief het waarborgen van vertrouwelijkheid, onafhankelijkheid en integriteit.
- *Praktische vaardigheden*: kunnen werken met AI-tools, output kritisch beoordelen, en waar nodig aanvullende maatregelen nemen zoals het uitvoeren van een Data Protection Impact Assessment of het opstellen van een intern AI-protocol.