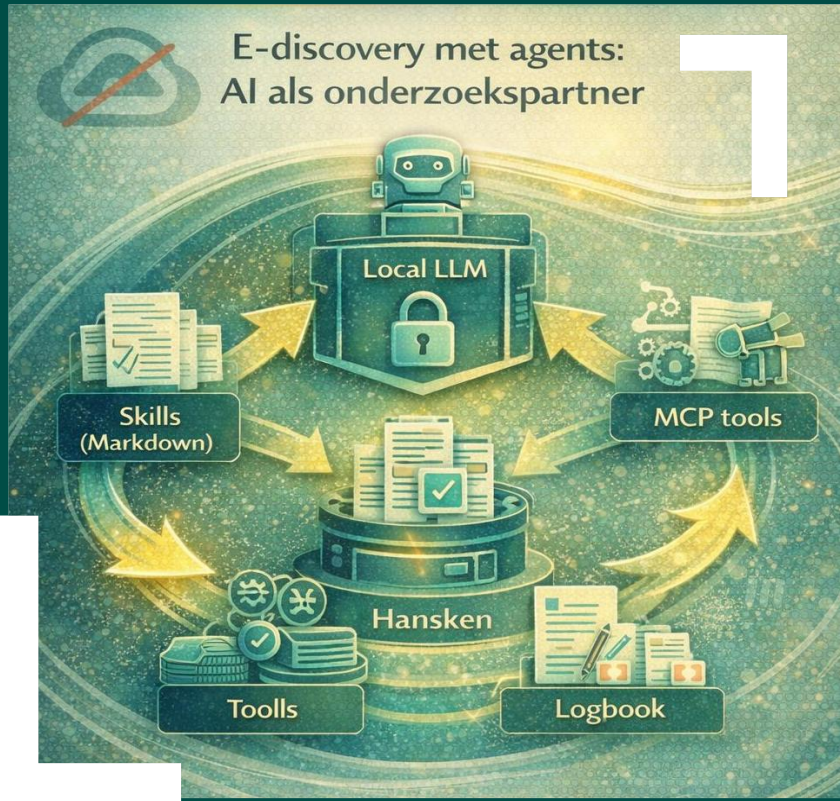
Hogeschool Leiden

E-discovery met agents

AI als onderzoekspartner

Hans Henseler,
Lector Digital Forensics & E-Discovery
Senior wetenschappelijk onderzoeker NFI

E-Discovery Symposium 2026



Hogeschool Leiden

Recente ontwikkelingen

Maart 2025

- Gemini 2.5 Pro (Google DeepMind)
- QwQ-Max (visual reasoning) (Alibaba)
- Gemma 3 4|12|27b (Google)
- OLMo 2, 32bm AI2

April 2025

- Llama 4 (Scout, Maverick, Behemoth) (Meta)
- GPT-4.1 (OpenAI)
- GPT-o3 & GPT-o4-mini (OpenAI)
- Phi-4-reasoning & Phi-4-reasoning-plus (Microsoft)
- Qwen 3 (Alibaba),
- DeepSeek Prover-V2 671B (DeepSeek)

Mei 2025

- Mistral Medium 3 (Mistral),
- GPT-4.1 mini (OpenAI)
- Falcon Arabic & Falcon H1 (ATRC — UAE)
- Claude 4 (Anthropic)

Juni 2025

- Magistral (first reasoning family) — Small (open) & Medium (API) (Mistral).
- Gemini 2.5 Pro & 2.5 Flash — GA (Google; adaptive “thinking” and fast/cheap tier).
- Gemini Robotics On-Device (Vision-Language-Action model) (Google DeepMind; on Pixel/Vision Pro-class devices).
- ERNIE 4.5 (open-sourced family, 10 variants) (Baidu).

Juli 2025

- Grok-4 (xAI; new flagship across X and API; “Heavy” tier).
- Devstral Small & Medium 2507 (Mistral; agentic coding models).
- Kimi K2 (open-weights MoE agentic model) (Moonshot AI; 7/11 release).
- Qwen3 updates — Qwen3-Coder, Qwen3-235B, Qwen-MT (Alibaba).
- Command A Vision (open-weights multimodal) (Cohere; enterprise doc/vision).

Augustus 2025

- GPT-5 (OpenAI) — new frontier model.
- GPT-OSS:20b/120b (open-weights family) (OpenAI).
- Claude Opus 4.1 (Anthropic) — stronger coding/agentic upgrade.
- DeepSeek-V3.1 (DeepSeek) — hybrid “thinking / non-thinking” mode; faster agents.
- Mistral Medium 3.1 (Mistral) — frontier-class multimodal refresh.
- Command A Reasoning (Cohere) — enterprise reasoning model.
- Command A Translate (Cohere) — specialized translation LLM (open-weights research release).
- Gemini 2.5 Flash Image (Google) — image gen/edit model under Gemini 2.5.
- NVIDIA Nemotron Nano 9B v2 (open) & Jet-Nemotron (2B/4B research) — hybrid Transformer/Mamba for fast reasoning on modest hardware.

September 2025

- Qwen3-Max-Preview (~1T parameters, preview access) (Alibaba/Qwen),
- Qwen3-VL (new vision-language line)
- Qwen3-Omni (omni-modal: text-image-audio-video; real-time)
- Claude Sonnet 4.5 — agent/computer-use focused upgrade (Anthropic)

Oktober 2025

- Claude Haiku 4.5 — small, very fast model (Anthropic)
- Gemini 2.5 Flash Image & Computer Use (Google).
- Qwen3-VL 2B & 32B releases (extra varianten)

November 2025

- Gemini 3 (Google) – nieuwe generatie (Pro) in app, AI Studio & Vertex.
- Grok 4.1 (xAI) – update met betere reasoning/multimodaal.
- Claude 4.5 Opus (Anthropic).
- ERNIE 5.0 (Baidu) – natively omni-modal.
- OLMo 3, A2I

December 2025

- DeepSeek V3.2 (DeepSeek).
- Mistral 3 familie + Mistral Large 3 (MoE) aangekondigd; Ministral 3 14B (instruct)
- ChatGPT 5.2 (OpenAI)

Januari 2026

- NVIDIA: nieuwe open modellen, data en tools (Nemotron-ecosysteem

Februari 2026

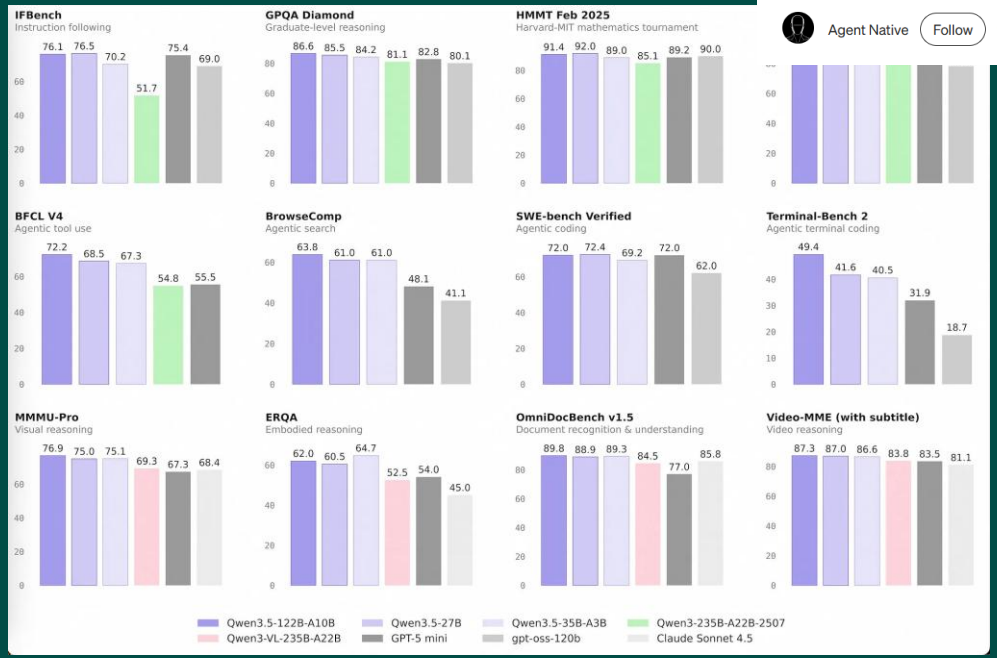
- OpenAI GPT-5.3-Codex (agentic coding model)
- Anthropic Claude Opus 4.6 + Sonnet 4.6 (incl. 1M token context “beta”)
- Google Gemini 3 Deep Think upgrade en Gemini 3.1 Pro
- Alibaba Qwen3.5-397B-A17B (MoE; 397B totaal, 17B actief per token), Qwen3.5-122B-A10B (MoE; 122B totaal, 10B actief) + Qwen3.5-35B-A3B (MoE; 35B totaal, 3B actief) + Qwen3.5-27B (dense)

Maart 2026

- Alibaba Qwen3.5-9B, 4B, 2B, 0.8B (dense “small” serie)
- *OpenAI GPT-5.4 (OpenAI) — nieuw frontier model (API + ChatGPT + Codex)GPT-5.4 Thinking (OpenAI) — reasoning-variant in ChatGPTGPT-5.4 Pro (OpenAI) — zwaardere compute variant via API*

Hogeschool Leiden

Qwen 3.5 …



Qwen 3.5 35B-A3B: Why Your $800 GPU Just Became a Frontier Class AI Workstation

Agent Native · Follow · 10 min read · Mar 1, 2026

Het dilemma van de moderne onderzoeker

Een (digitale) berg aan bewijs

- Quan en Joseph worden aangehouden met een koffer van Quan die veel cash geld bevat.
- Het bewijs: digitaal beslag van 10 gegevensdragers:
  - 7 Mobiele telefoons
  - 1 MacBook
  - 1 Windows Laptop
  - 1 USB Drive

Tegenstrijdige verklaringen

- Verklaring van Quan: “Het geld is een gift van mijn overladen grootvader om mijn studie aan de universiteit te betalen en om mijn gokschulden af te lossen.”
- Verklaring van Joseph: “Ik heb gewoon een taxi en verhuurbedrijf. Ik ben ingehuurd om iemand bij het vliegvel op te halen. Ik weet niets van het cash geld."

Vragen van de onderzoeker

- Kloppen de verklaringen van Quan en Joseph?
- Zijn er tegenstrijdigheden tussen hun verklaringen en het digital bewijs?
- Hoe kun je verbanden vinden tussen Quan, Joseph en de misterieuze Nerijus in duizenden chatberichten.
- Waar komt het cash geld nu echt vandaag?

- Grote hoeveelheden chat- en telefoongegevens (EncroChat/Sky ECC).
- Beperkte tijd en capaciteit bij politie/FO om alles handmatig te lezen.
- Vraag: hoe houd je het onderzoek én uitlegbaar én controleerbaar voor de rechter?

Hogeschool Leiden

Een raamwerk voor AI-taken in de Crystal Clear-zaak

Reduce (essentie destilleren)

- **Generieke functie:** "Van veel tekst naar een korte, gerichte samenvatting."
- **Toepassing in Crystal Clear:** *"In plaats van duizenden WhatsApp-berichten zelf te lezen, vraag je de AI om alle gesprekken tussen Joseph en Nerijus samen te vatten. Je hebt de belangrijkste onderwerpen in enkele minuten in beeld, in plaats van na dagen lezen."*

Transform (vorm veranderen)

- **Generieke functie:** "De vorm veranderen, maar niet de kern van de betekenis."
- **Toepassing in Crystal Clear:** *"Je hebt het vermoeden dat Quan en Joseph elkaar vóór de aanhouding hebben ontmoet. Je vraagt de AI: 'Maak een tijdlijn van alle locatiegegevens van Quan en Joseph in de 48 uur vóór hun aanhouding.' De AI zet ruwe GPS-logs om in een eenvoudige, visuele tijdlijn."*

Generate (uitbreiden en creëren))

- **Generieke functie:** "Van een kleine prompt naar een nieuwe, uitgebreide tekst."
- **Toepassing in Crystal Clear:** *"Je hebt drie kernberichten en een verdachte locatie gevonden. Je vraagt de AI: 'Schrijf een voorlopig rapport dat deze bewijselementen met elkaar verbindt en doe een hypothese over de herkomst van het geld.' De AI maakt de eerste concepttekst en bespaart je uren schrijftijd.*

Hogeschool Leiden

Drie jaar geleden leek dit ver weg

2023

- LLM's = vooral samenvatten / chatten
- 'Agent die forensische vragen oplost met tool chaining' klonk als toekomstmuziek

Nu in 2026

Een CLI-agent (bijv Codex) kan:

- Project selecteren → data inventariseren
- Multi-tool queries → bewijs disciplineren
- Rapport + visualisaties genereren

De doorbraak: MCP server + tool catalog + skills + repo-structuur

Hogeschool Leiden

Ons doel: een copilot voor de onderzoeker

Onze aanpak: copilot

- **Rol:** ondersteunt de expert, vervangt hem/haar niet.
- **Functie:** vat data samen, doet suggesties voor verbanden, automatiseert repeterende taken.
- **Besluitvorming:** de (digitale) onderzoeker neemt alle uiteindelijke, cruciale beslissingen.
- **Basisprincipe:** human in the loop – de mens blijft aan het stuur.

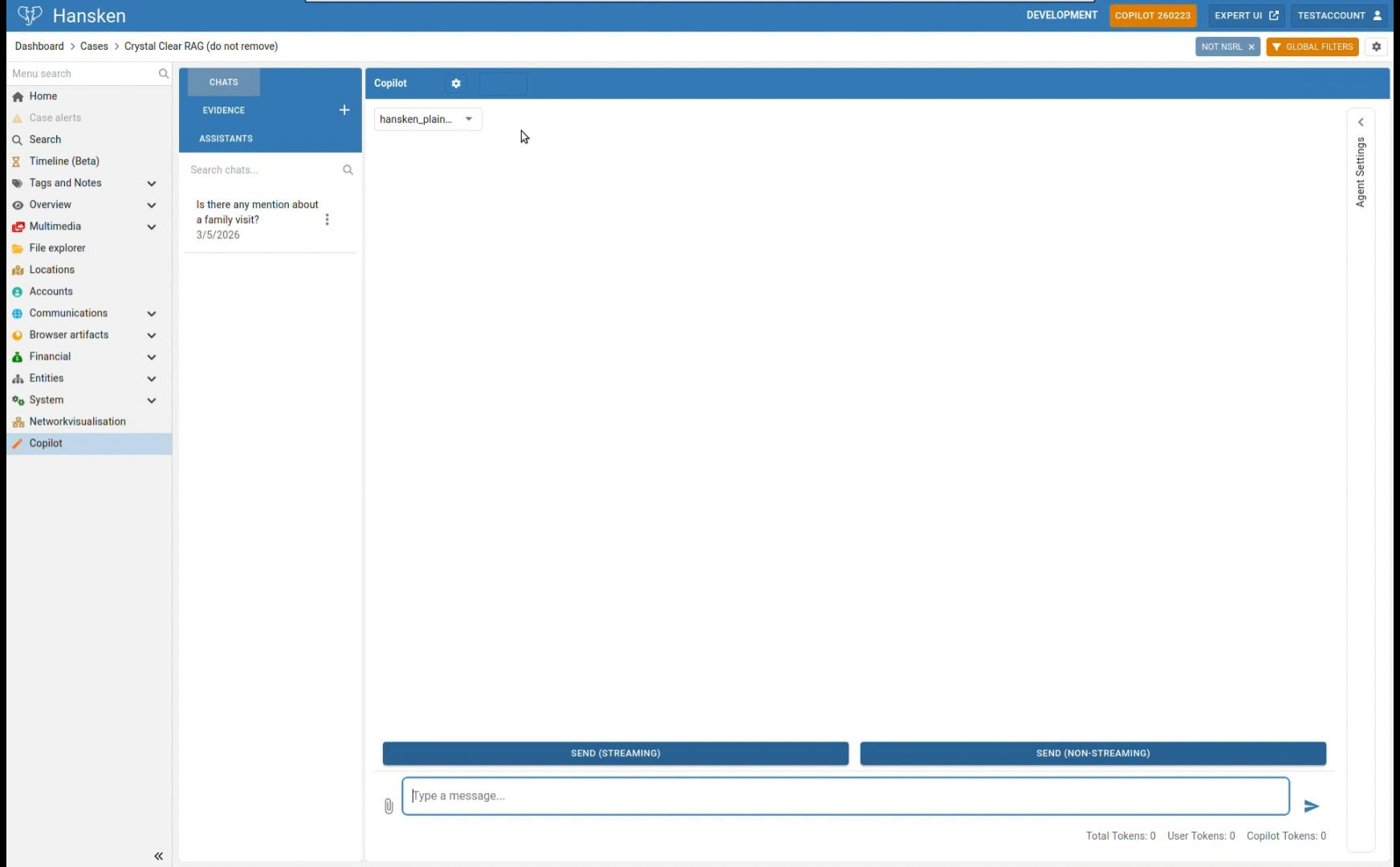
Dus geen: autopilot

- **Rol:** vervangt de expert.
- **Functie:** werkt volledig autonoom.
- **Besluitvorming:** de AI neemt zelfstandig beslissingen.
- **Basisprincipe:** forensisch en ethisch onwenselijk voor complexe onderzoeken.

Hogeschool Leiden

Voorbeeld van RAG Vraag & Antwoord

- **Vraag:**
  - Worden er naast euro's ook andere betaalmiddelen gebruikt?
- **Copilot antwoord:**
  - Ja. Een contactpersoon genaamd XXXXXXXX bespreekt met de verdachte een nieuwe regeling voor een "club in Ibiza" en stelt expliciet: "Delivered drugs will be paid in Bitcoins.....".
- **Spoor hint:**
  - Evidence Trace ID: 6c5099cb-f06f-40ca-8051-57f392137ed4:0-0-0-1-0-1-2-86-2-3-0.


Hansken
DEVELOPMENT
COPILOT 260223
EXPERT UI
TESTACCOUNT
Dashboard > Cases > Crystal Clear RAG (do not remove)
NOT NSRL
GLOBAL FILTERS
Menu search
Home
Case alerts
Search
Timeline (Beta)
Tags and Notes
Overview
Multimedia
File explorer
Locations
Accounts
Communications
Browser artifacts
Financial
Entities
System
Networkvisualisation
Copilot
CHATS
EVIDENCE
ASSISTANTS
Search chats...
Is there any mention about a family visit?
3/5/2026
Copilot
hansken_plain...
Agent Settings
SEND (STREAMING)
SEND (NON-STREAMING)
Type a message...
Total Tokens: 0
User Tokens: 0
Copilot Tokens: 0

Hogeschool Leiden

Uitdagingen voor een lokale AI-copilot

Het contextprobleem (hoeveel kan het model tegelijk zien?)

- Ons lokale model heeft een beperkt contextvenster (bijv. 32k tokens). Het kan steeds maar een klein deel van de zaakdata tegelijk bekijken.
- Het kan het dossier van 250k+ tokens niet in één keer analyseren. We hebben dus een manier nodig om grote datavolumes beheersbaar te maken.

Het koppelen van entiteiten (entity resolution)

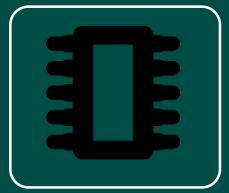
- Forensische data is gefragmenteerd. Eén persoon, “Joseph Prinse”, komt in verschillende sporen voor als telefoonnummer, WhatsApp-ID en e-mailadres.
- Het model weet niet vanzelf dat al deze identificatoren bij dezelfde persoon horen en kan ze zonder hulp niet koppelen.

Het redeneerprobleem (beperkt onderzoeksplan)

- Ons lokale model (gpt-oss-20b) kan goed instructies opvolgen, maar kan niet zelfstandig een complexe onderzoeksstrategie in meerdere stappen ontwerpen, zoals sommige grote frontier-modellen dat kunnen.
- Het moet worden gestuurd door een onderzoeksplan van de expert.

Probleem: hoe ontwerpen we een copilot op basis een lokaal werkend LLM?

Wat is een agent? Drie kernvaardigheden



Geheugen:

- Kan eerdere interacties en resultaten “onthouden” en die gebruiken voor de volgende stappen.
- Cruciaal bij onderzoek dat uit meerdere stappen bestaat.

Plannen & Beslissen:

- Kan een complex doel opdelen in kleinere deeltaken.
- Kan redeneren, zichzelf bijsturen en bepalen wat de volgende stap moet zijn.

Tools aanroepen:

- Dit is het echte verschil met een gewone chatbot.
- De agent kan zelf externe tools kiezen en gebruiken (bijv. onze PersonLinker) om informatie op te halen die hij zelf niet heeft

- Een agent beantwoordt niet alleen een vraag, maar maakt een plan en verzamelt zelf de middelen die nodig zijn om die vraag goed te beantwoorden.

Hogeschool Leiden

Tools en het Model Context Protocol (MCP)

MCP = Model Context Protocol — een open standaard (Anthropic) waarmee een AI-agent tools kan aanroepen

Metafoor: MCP is de universele stekker tussen de LLM en Hansken

Waarom MCP en niet gewoon een API-wrapper?

- Gestandaardiseerd — werkt met elke MCP-compatibele client
- Tool discovery: agent ontdekt zelf welke tools beschikbaar zijn
- Veilig: server bepaalt wat wel/niet mag

Hogeschool Leiden

Tool catalog als agent-handleiding

- Tool catalog: de MCP server biedt tool-beschrijvingen aan die de agent raadpleegt
  - Agent kan toolkeuze maken (hanskenSearchByProperty vs hanskenAdvancedSearch)
  - Agent kan anti-patterns vermijden (catalog waarschuwt expliciet)
  - Agent kan tool chaining plannen op basis van WHEN TO USE beschrijvingen
- Concreet voorbeeld uit de catalog:

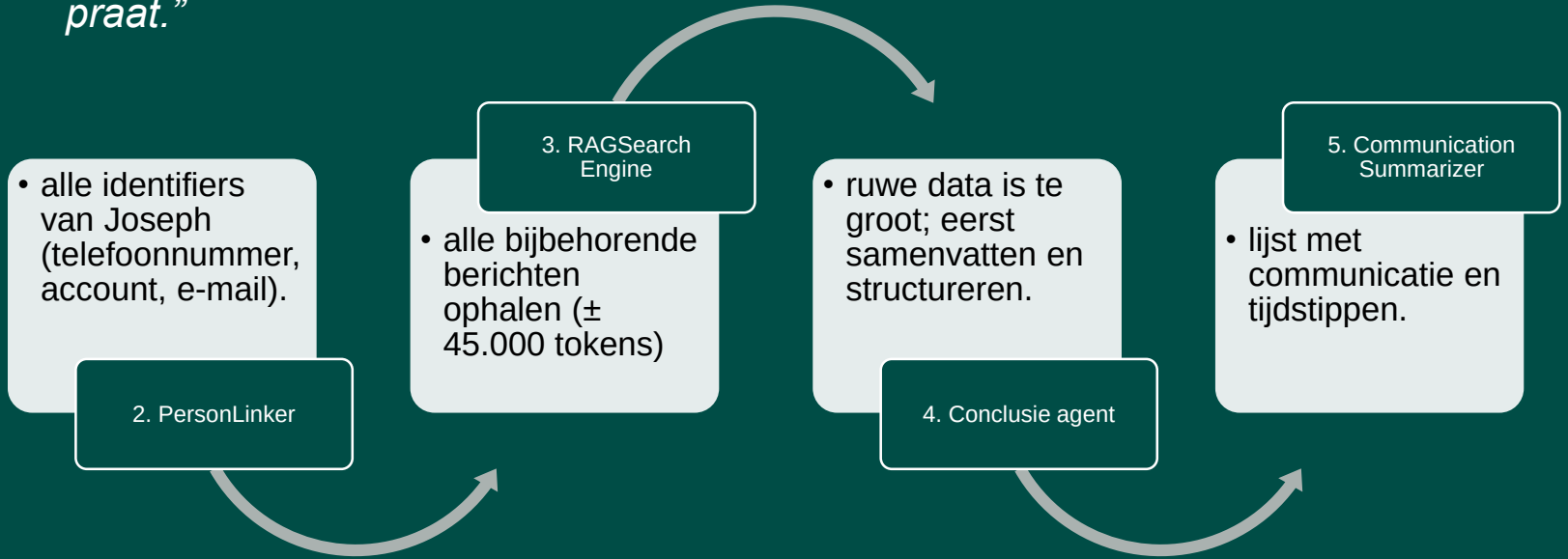
Tool	Doel	Let op
hanskenFindAccounts	Ownership hints	Kan noisy zijn
hanskenAdvancedSearch	Multi-conditions met HQL	Krachtigste tool
hanskenGetTraceDetails	Metadata (IDs/phones)	Zonder body
hanskenGetTraceTextContent	Body/content ophalen	Duur, selectief gebruiken

Hogeschool Leiden

Agentic RAG in de praktijk (Deel 1)

Van vraag naar bewerkte data

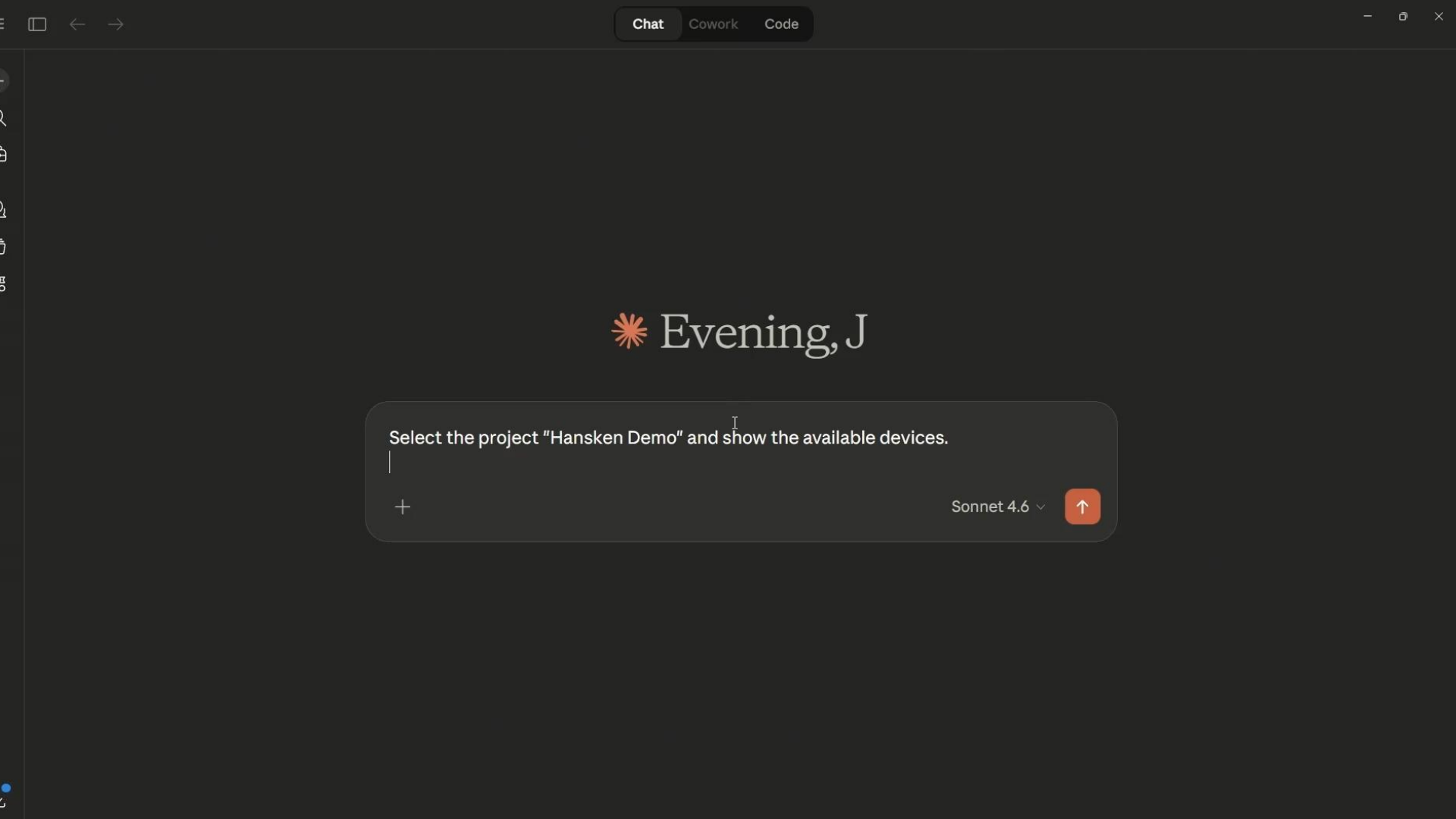
1. Vraag: *“Vind alle communicatie van Joseph Prinse en laat zien met wie hij het meest praat.”*

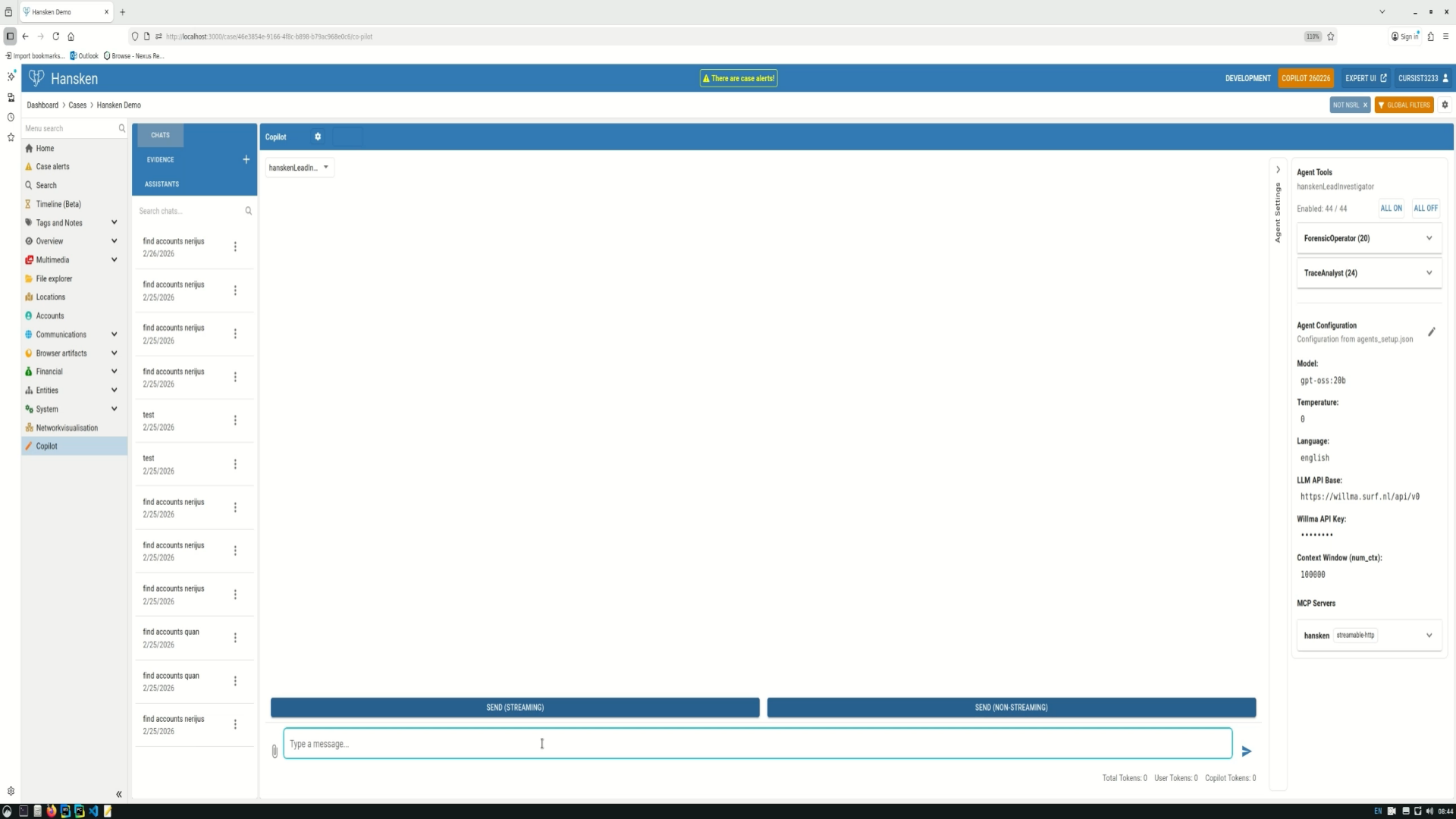


Agentic RAG in de praktijk (Deel 2)

Van bewerkte data naar rapport

Elke toolstap en tussenuitkomst kan worden gelogd, zodat de hele keten desnoods in de rechtszaal kan worden gereconstrueerd.


Chat
Cowork
Code
Evening, J
Select the project "Hansken Demo" and show the available devices.
Sonnet 4.6


Hansken Demo
http://localhost:3000/case/46e3854e-9166-4f8c-b898-b79ac968e0c6/co-pilot
Import bookmarks...
Outlook
Browse - Nexus Re...
Hansken
There are case alerts!
DEVELOPMENT
COPILOT 260226
EXPERT UI
CURSIST3233
Dashboard > Cases > Hansken Demo
NOT NSRL
GLOBAL FILTERS
Menu search
Home
Case alerts
Search
Timeline (Beta)
Tags and Notes
Overview
Multimedia
File explorer
Locations
Accounts
Communications
Browser artifacts
Financial
Entities
System
Networkvisualisation
Copilot
CHATS
EVIDENCE
ASSISTANTS
Search chats...
find accounts nerijus 2/26/2026
find accounts nerijus 2/25/2026
find accounts nerijus 2/25/2026
find accounts nerijus 2/25/2026
test 2/25/2026
test 2/25/2026
find accounts nerijus 2/25/2026
find accounts nerijus 2/25/2026
find accounts nerijus 2/25/2026
find accounts quan 2/25/2026
find accounts quan 2/25/2026
find accounts nerijus 2/25/2026
Copilot
hanskenLeadIn...
SEND (STREAMING)
SEND (NON-STREAMING)
Type a message...
Total Tokens: 0
User Tokens: 0
Copilot Tokens: 0
Agent Settings
Agent Tools
hanskenLeadInvestigator
Enabled: 44 / 44
ALL ON
ALL OFF
ForensicOperator (20)
TraceAnalyst (24)
Agent Configuration
Configuration from agents_setup.json
Model:
gpt-oss:20b
Temperature:
0
Language:
english
LLM API Base:
https://willma.surf.nl/api/v0
Willma API Key:
••••••••
Context Window (num_ctx):
100000
MCP Servers
hansken
streamable-http

Hogeschool Leiden

Reality check: agents zijn geen wondermiddel

- Agents lijken de volgende stap in digitaal onderzoek, maar ze lossen niet alles op.
- Zeker in de forensische praktijk, waar veel op het spel staat, brengen ze nieuwe risico's en aandachtspunten mee die je actief moet beheersen.

Betrouwbaarheid & kwetsbaarheid:

- Agents kunnen nog steeds "stuk gaan". In een complex meerstappenplan kan het mislopen bij stap 3, in stap 4 de verkeerde tool worden gekozen of kan het proces stilvallen. Dat vraagt om robuuste foutafhandeling, bewaking en waar nodig menselijke tussenkomst.

Tijd & rekencapaciteit:

- Elke "gedachte" of tussenstap betekent vaak een nieuwe LLM-aanroep. Dat kost tijd en rekenkracht. Voor sommige vragen is een eenvoudigere werkwijze , bijvoorbeeld één gerichte zoekactie of een simpele RAG-query, sneller en efficiënter dan een complete Agentic AI-opzet.

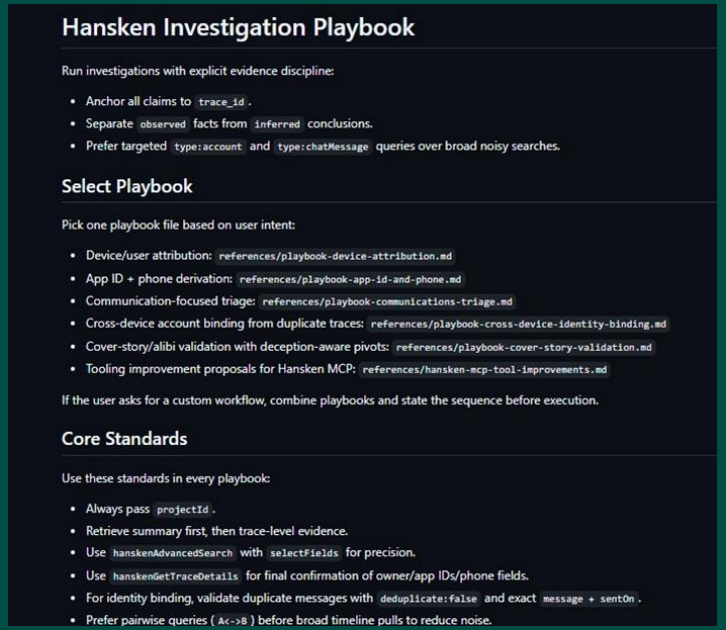
Traceerbaarheid & controleerbaarheid:

- Hoe meer stappen en zelfcorrecties, hoe lastiger het wordt om te zien hoe een conclusie precies tot stand is gekomen. "De agent besloot dit" is geen toereikende uitleg.
- Het ontwerp moet voorzien in een gedetailleerde logging, zodat achteraf te reconstrueren is welke prompts, tools en data tot welke tussenstappen en conclusies hebben geleid.

Hogeschool Leiden

Over LLM Skills en Playbooks

- **Skills** – Herbruikbare, modulaire acties die een AI-agent kan uitvoeren, zoals het ophalen van data, het aanroepen van een API of het verwerken van een document. Skills vormen de bouwstenen van een agent.
- **Playbooks** – Voorgedefinieerde workflows die beschrijven hoe en wanneer een agent bepaalde Skills inzet om een doel te bereiken. Ze bieden structuur en consistentie in complexe, meerstaps-processen.



Hansken Investigation Playbook

Run investigations with explicit evidence discipline:

- Anchor all claims to `trace_id`.
- Separate `observed` facts from `inferred` conclusions.
- Prefer targeted `type:account` and `type:chatMessage` queries over broad noisy searches.

Select Playbook

Pick one playbook file based on user intent:

- Device/user attribution: `references/playbook-device-attribution.md`
- App ID + phone derivation: `references/playbook-app-id-and-phone.md`
- Communication-focused triage: `references/playbook-communications-triage.md`
- Cross-device account binding from duplicate traces: `references/playbook-cross-device-identity-binding.md`
- Cover-story/alibi validation with deception-aware pivots: `references/playbook-cover-story-validation.md`
- Tooling improvement proposals for Hansken MCP: `references/hansken-mcp-tool-improvements.md`

If the user asks for a custom workflow, combine playbooks and state the sequence before execution.

Core Standards

Use these standards in every playbook:

- Always pass `projectId`.
- Retrieve summary first, then trace-level evidence.
- Use `hanskenAdvancedSearch` with `selectFields` for precision.
- Use `hanskenGetTraceDetails` for final confirmation of owner/app IDs/phone fields.
- For identity binding, validate duplicate messages with `deduplicate:false` and exact `message + sentOn`.
- Prefer pairwise queries (`A<->B`) before broad timeline pulls to reduce noise.

Hogeschool Leiden

Een hint is nog geen bewijs, maar kan wel tot bewijs leiden

Gubanov vs. Brignoni (zie discussie op LinkedIn hier)

De LLM als ‘digitale speurneus: Geeft aanwijzingen, geen bewijs. **(visie van Gubanov)**

Voorbeeld: “Dit chatgesprek wijst op een overdracht.”

Voorbeeld: “Kijk in logbestand X rond tijdstip Y.”

De uitdaging van herleidbaarheid: Het pad naar de hint moet reproduceerbaar en methodologisch verantwoord zijn (zorg van Brignoni, verwijzend naar de Daubert-standaard).

Ons uitgangspunt: “Een door AI gegeven hint moet altijd zelfstandig met reguliere forensische methoden worden getoetst; het uiteindelijke bewijs berust op de uitkomst van die toetsing, niet op de hint zelf.”

Hogeschool Leiden

Geen Daubert in Nederland, wél stevige kwaliteitseisen

- **Key Principles for Admissibility:**
  - **Transparantie:** Het gebruik van het LLM moet worden vastgelegd (bijvoorbeeld in logbestanden). We moeten kunnen uitleggen dát en op welke wijze een tool is gebruikt om op een spoor te komen.
  - **Herleidbaarheid:** Het uiteindelijk gerapporteerde bewijsmateriaal moet via reguliere forensische methoden zijn terug te voeren op de brondata, los van de oorspronkelijke AI-hint.
  - **Reproduceerbaarheid:** Een andere deskundige moet, wanneer hij dezelfde methode op dezelfde data toepast, in wezen tot dezelfde bevindingen over het bewijs kunnen komen.
- **Conclusie:**
  - *“De juridische toets richt zich niet op de interne werking van het LLM, maar op de kwaliteit en transparantie van het forensische onderzoek dat op de hint volgt.”*

Hogeschool Leiden

Richtlijnen voor het gebruik van LLM’s in forensisch onderzoek

Gebruik alleen als ondersteuning, niet als primaire bron van bewijs.

Valideer elke hint zelfstandig met reguliere forensische methoden.

Leg het gebruik vast (prompts, modelversie, parameters, tijdstip, enz.).

Zorg voor herleidbaarheid; gebruik je RAG, verwijs dan naar de geraadpleegde bronnen

Rapporteer alleen controleerbare bevindingen, niet de losse suggesties van de AI.

Wees voorzichtig met ‘niet gevonden’; de afwezigheid van een hit is zelden op zichzelf overtuigend bewijs.

Hogeschool Leiden

Conclusies: RAG en de rolverdeling mens–AI

- **RAG = slimme zoek- en samenvattool**

  Zoekt en vat samen op basis van de stukken die je aanlevert; een krachtige tekstsamenvatter, maar passief.

- **Agentic RAG = onderzoeksassistent**

  Volgt een expliciet stappenplan (bijv. de 7 stappen), schakelt tussen meerdere tools (zoeken, samenvatten, tijdlijn, netwerk) en levert conceptanalyses – geen eindbewijs.

- **De digitaal onderzoeker bepaalt de koers**

  Formuleert de onderzoeksvraag, kiest de methode, beoordeelt de data en schrijft het rapport. Hij/zij blijft verantwoordelijk voor de conclusies richting rechtbank.

- **De AI-copiloot doet het zware werk**

  Helpt bij zoeken, filteren en structureren van grote datastromen. Elke door de AI gevonden hint moet herleidbaar zijn in het dossier en door de onderzoeker worden gevalideerd voordat zij als bewijs wordt gebruikt.

Hogeschool Leiden

Expertise en Recht

- Redactioneel samen met Diederik Aben
- LLM’s in forensisch onderzoek: hints, herleidbaarheid en juridische toetsing
- 2025-5 October 2025

Hogeschool Leiden

Bedankt voor jullie aandacht!

henseler.h@hsleiden.nl

https://www.linkedin.com/in/henseler/
https://www.hsleiden.nl/digital-forensics