Opdrachtomschrijving onderzoeksproject Casper van Wetten



Recentelijk heeft Wouter Westerkamp zijn onderzoek naar het automatisch genereren van proces-verbaal (pv) bij de politie afgerond. Uit dit onderzoek blijkt, onder andere, dat de ontwikkeling en implementatie van LLM systeem om op basis van het verhoor automatisch pv’s te generen voor specifieke delicten waarschijnlijk mogelijk is. Verder is het duidelijk dat er meer onderzoek vereist is voordat een dergelijk systeem bruikbaar kan zijn in de dagelijkse uitvoering.

Nu Wouter heeft aangetoond dat een dergelijk systeem een pv kan opzetten wat voldoet aan de inhoudsvereisten (29a lid 1 Sv) is het belangrijk om te onderzoeken of een dergelijk systeem een pv kan afleveren wat omkleed is met de volledige vereisten der wet. Is het mogelijk voor een LLM-gebaseerd systeem om een verklaring “zo veel mogelijk in zijn eigen woorden” (29a lid 2 Sv) op te nemen? In Nederland zijn pv’s vaak samenvattingen, geselecteerd op basis van relevantie voor het verdere strafproces [1], kan een dergelijk systeem elke keer correct de relevante informatie selecteren?  
Naast de juridische pijnpunten zijn er ook een aantal technische. Is het mogelijk om het hallucineren van de gebruikte LLM volledig tegen te gaan? Hoe traceerbaar en uitlegbaar zijn een dergelijk systeem kunnen zijn? Is het mogelijk om een white-box model te implementeren i.p.v. de gebruikelijke black-box modellen? Hoe kan een dergelijk systeem voldoen aan de data-veiligheid en privacy vereisten voor de zeer gevoelige informatie waar zij mee zou werken? Wat voor trainingsdata zou er gebruikt kunnen worden voor dergelijk systeem? Hoe aanpasbaar moet het zijn indien er nieuwe regelgeving ingevoerd wordt?

In mijn onderzoek zou ik graag het stokje van Wouter overpakken en (een aantal van) deze details nader bekijken. Specifiek ben ik benieuwd naar de mogelijkheid of een dergelijk systeem een met de vereisten der wet omklede pv kan genereren en naar hoe dit realistisch op in-house geïmplementeerd kan worden.

notes:

en aantal Kan het gebruik van een robot wat de transcriptie samenvat tot een p-v analogisch gezien worden aan een administratief medewerker die werk uit handen neemt van de verbalisant? Tot hoeverre kan de hallucinatie van een transformermodel vermindert worden, en op welk punt kan zij gelijkgesteld worden met de menselijke fouten van een verbalisant? Als de transcriptie (ofwel de input voor het model) en de proces-verbaal (ofwel de output van het model) deel blijven van de gehele rechtsketen, voldoet dat aan de integriteitsplicht, met name de noodzaak aan openheid en transparantie van totstandkoming van het document? Wat als de transformer gepaard wordt met een (white- of black-box) explainer model?

Veelal van deze vragen vallen onder het overkoepelende vraagstuk van wat er precies

Het is zeker mogelijk om een lokaal-runnend transformer model te ontwerpen, maar kan deze ook voldoen aan de functionele vereisten en presteren tot (tenminste) het niveau van het PoC? Desondanks dat deze waarschijnlijk gebruikmaakt van veel krachtigere computers, grotere datasets en meer parameters? Kan een transformer model, wat bij ontwerp een blackbox is, voldoen aan de vereisten van openheid en transparantie w.b. de dataverwerking? Kan een transformer model, met haar hallucinatieproblemen, altijd vertrouwd worden dat zij voldoet aan juistheid en volledigheid? Kan zij überhaupt vertrouwd worden om te voldoen aan juistheid en volledigheid? Wat voor rol kan menselijke controle spelen om toch aan deze standaarden te voldoen?

(Een deel van) Dit onderzoek betreft het verkennen van deze problemen en het uitdiepen van mogelijke oplossingen.

Bronnen

[1] M.L. Komter, Delikt en Delinkwent, *De verschriftelijking van verdachtenverhoren,* 18-12-2019