

Henk Hutink: Een mail van een gerenommeerde AI specialist.

Ik stuur jullie deze mail om een indruk te geven van de problemen om medische data geschikt te maken voor artificial intelligence in Nederland. Ondanks het feit dat Europa dezelfde GDPR wetgeving heeft, is databeschikbaarheid in elk land anders geregeld. Dit nog los van het feit hoe databeschikbaarheid in de Verenigde Staten en China georganiseerd is. In verband met privacy redenen heb ik aanknopingspunten naar de herleidbaarheid van de zender mailschrijver zwart gemaakt.

=====

Dag Henk Hutink

Momenteel heb ik vanuit het XXX -project in het XXX -Ziekenhuis actief een samenwerking opgestart met de het Universitair ziekenhuis van XXX in Duitsland, waarbij we een federated learning techniek willen toepassen voor het trainen van een algoritme met lokale data. Het algoritme hebben we samen met de Universiteit XXX ontwikkeld, echter we hebben een tekort aan data om het te testen. In XXX zijn deze beschikbaar, echter het delen van data vanuit Duitsland is zeer omslachtig en moeilijk. O.b.v. federated learning proberen we ook een elegante oplossing te vinden voor de bestaande administratieve en legale hindernissen.

Het beschikbaar stellen en uitwisselbaar maken van data vanuit ziekenhuizen is een issue dat hoge prioriteit verdient, er is immers een groot gebrek aan kwaliteitsvolle data. Vanuit het XXX-Covid project , waaraan ook het XXX meewerkt, zijn we erin geslaagd om 4000 CT scans te verzamelen uit 7 Europese landen, echter de contractuele en legale beslommeringen die hiermee gepaard gingen werkten een enorm vertragend - zelfs in een crisissituatie zoals een pandemie. Het verzamelen van beelden en data ging zeer moeizaam bij veel ziekenhuizen, terwijl nochtans het doel van overstijgend gemeenschappelijk belang is. De Nederlandse ziekenhuizen waarmee we samenwerkten hadden een zeer wisselende aanpak, wat aangeeft dat er in praktijk nog veel vragen en onzekerheden bestaan over de manier waarop met dit soort data wordt omgesprongen, zelfs voor researchdoeleinden. Dit is vooral het geval bij de niet-academische centra. Er is zeker behoefte aan meer eenduidigheid en richtlijnen qua aanpak. We zijn met dit project gestart in maart, en konden het finale algoritme slechts trainen en testen in december 2020. Wat de meeste tijd heeft gekost (6-7 maanden) zijn

- de contractuele en legale hindernissen en onduidelijkheden
- het voorbereiden van de data: gebrek aan uniformiteit (pseudonimisatie) en middelen
- het doorsturen en ontvangen van de data (noodzaak tot gemeenschappelijke gateway c.q. platformen)

vriendelijke groet

██████████

=====