AI – hur säkrar vi indata?

Rikard Edgren, Nordic Medtest

rikard.edgren@nordicmedtest.se

”Artificiell Intelligens”

- AI kan betyda många saker
  - Generell intelligens som kan ta över världen
  - ”Narrow AI” som lär sig bli oerhört bra på ett specifikt område
  - En väldigt avancerad algoritm

- I denna presentation utgår jag från att AI är programvara som förändras (till det bättre?) baserat på ny indata

Barnsjukdomar

- Tay – Chatbot på Twitter
  https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Databaserad Machine Learning

- Rå data (unsupervised)
  - Exempelvis data som användare förser systemet med
- Kurerad data (supervised)
  - Exempelvis den grunddata som systemet laddats med
- Förstärkning av data (reinforcement)
  - Data som gav bra resultat värderas högre
- Data från andra områden (transfer learning)
  - Exempelvis att lära sig svenska från andra områden än vården

Olika typer av data

- strukturerad vs. ostrukturerad data
- strukturerbar vs. ostrukturerbar data
- konstruerad vs. verklig data
- kvantitet vs. kvalitet på data
- generell vs branschspecifik data
- hur ser indata ut i framtiden? kan vi veta det?

Risk – för lite data

- Testa systemet för att se vad konsekvenserna blir
- Granskning av metoderna för att få mer data
- Kritisk granskning av eventuella lösningar för att köra på ändå

Risk – icke-representativ data

- Mångfald hos de som arbetar med systemet
- Titta med många perspektiv
- Titta på aggregerade vyer
  - Kräver strukturerad data
  - Kan visa på snedvridningar i datastrukturer
- Kritisk granskning av eventuella lösningar för att ”rätta till”

Risk –sabotage

- Övervakning av indata
- Svartlistningar
  - Fula ord, höga doseringar m.m.
- Wallraffa?
- Säkerhetstestning

Risk – dålig data, skräp och brus

- Inbyggda varningssystem
- Strukturering av data
  - Ex. poängsättning av datakvalitet
- Titta på aggregerade vyer
- Kvalitativa stickprov

Risk – buggar

- ”Vanlig” testning
  - Förstå systemet, och eventuella speciallösningar
  - Utmana systemet, och kraven
  - Testa realistiskt, och orealistiskt
- Förutom funktionalitet, utvärdera
  - Prestanda
  - Stabilitet
  - Säkerhet
  - Användarvänlighet

Risk – etiska gråzoner

- Bör patienten godkänna AI:s medverkan innan indata behandlas?
- Får/Bör vi lagra patientens indata?
- Bör en människa kopplas in vid viss indata?
- Bör AI märka om en patient vill ha läkemedel av fel anledning?
- Vården kan bli sämre för de som har svårt att formulera sina besvär på ett sätt som datorn kan tolka, är det OK?
- Bör AI kunna hitta på egna åtgärder?

Kvalitetssäkring av indata

- Använd människor och maskiner
- En mångfacetterad teststrategi kan fånga oväntade sidoeffekter
- Var inte rädd för att testa även om det inte finns något facit
- Se till att många personer tänker test & kvalitetssäkring

Utdata…

- Är ju det allra viktigaste, det är ju resultatet av systemet
  - kan ha säkerhetsmekanismer som märker om det blir ”för” fel (”AI som kontrollerar AI”?)
  - bör övervakas även av människor (analyserbar beslutslogg)
  - vilka kan avgöra vad som är ett bra resultat?
  - får utdata användas som ny indata?

Frågor

- ???

Rikard Edgren

rikard.edgren@nordicmedtest.se


Träffa oss på
Mötesplats e-hälsa!
Monter: B04:12
Socialstyrelsen
eHälsomyndigheten
Sveriges Kommuner och Landsting
inera
tillsammans
för utveckling av
svensk e-hälsa
Tankar och medskick
Vad vill du lyfta? Skriv här!