

**Titel:** Hvor gode kan syntetiske sundhedsdata blive – og hvor bliver man forpligtet til at bruge dem?

**Oplægsholder:** Thor Hvidbak, Senior Manager, Deloitte Consulting

**Angiv tre keywords for oplægget:**

Syntetiske sundhedsdata, dataminimering, generativ AI

**Resumé:**

Syntetiske data er for alvor på vej frem på sundhedsområdet og i den bredere offentlige sektor. Men hvad bliver de brugt til, hvad kan de, hvad kan de ikke – og hvor er de mest interessante projekter?

Grundlæggende er der tre typer behov, som driver anvendelsen af syntetiske sundhedsdata, "de tre D'er": Datadeling, Datadiversitet (bedre træningsdata til AI-modeller) og Dataminimering. Syntetiske data kan skabes af konkurrerende, generative AI-netværk, og én af fordelene ved metoden er et langt bedre trade-off mellem anvendelighed og privatlivsbeskyttelse end med konventionelle anonymiseringsmetoder.

Med afsæt i konkrete projekter og eksempler på syntetiske data vil oplægget forholde sig til, hvor god en datakvalitet, der reelt kan opnås med syntetiske data. Og nok så vigtigt hvordan datakvaliteten kan varedokumenteres på en brugervenlig måde. Som noget helt nyt vil oplægget også beskrive en metode til tjekke hvor robust, en anonymisering er – en metode, der også kan bruges på andre typer anonymisering, hvilket vil give dataansvarlige myndigheder et bedre grundlag for at risikovurdere datasæt.

Generativ AI har med ChatGPT, DALL-E, Midjourney m.fl. fået sit folkelige gennembrud i 2023, men metoderne har faktisk længe været brugt til at generere syntetiske data – eksempelvis i Region Sjælland i form af syntetiske henvisningsdata til brug for AI-modeludvikling. Helt tilbage i januar 2018 anbefalede det norske datatilsyn myndigheder at bruge generative AI-netværk til at lave syntetiske data til test af it-løsninger. Efterfølgende har de tilmed udskrevet bøder til myndigheder, der alligevel har brugt personfølsomme produktionsdata til test og udvikling. Det samme kommer til at ske i resten af Europa. Syntetiske data har derudover potentiale til at accelerere AI-udviklingen ved at skabe benchmark-datasæt af høj kvalitet, som kan bruges til at sammenligne kvaliteten af algoritmer.

I april 2023 har Folketingets udvalg for hhv. Sundhed og Digitalisering & It selv eksplicit tematiseret potentialerne ved syntetiske data i deres oplæg til en fælles udvalgs høring om sundhedsdata, hvor oplægsholderen deltog.

**Kort præsentation af oplægsholder:** Thor har arbejdet med digital sundhed i 12 år, både som it-chef og som rådgiver, og han har ansvar for sundhedsdataområdet, der spænder fra klassiske analyseprojekter til AI-udvikling. Thor har tidligere haft forskellige it-lederstillinger i Region Hovedstaden, herunder som informationssikkerhedschef.