

Elefanten i rummet



ChatGPT versus Operational Excellence

**ChatGPT dræber innovation
og dumper på Privacy og Security**

**ChatGPT interviewes om Compliance
af en compliance expert**

Af Erik J. Andersen, Cand. Polit., CEO og Senior Lead Auditor i Symbic

Dette paper handler om AI og især OpenAI's internetbaserede ChatGPT standardtilbud

Introduktion

Explainer: Metakognition er menneskers bevidste viden om hvordan vore egne tankeprocesser fungerer, og deres stærke og svage sider. Det sætter os bla. i stand til at vælge strategier for indlæring og samspil med andres tanker, og følelser, selvom de ikke udtrykkes explicit.

ChatGpt 4 generationen er ikke Metakognition, og understøtter ikke Operational Excellence

Der er endnu nogle år til AI når et modent stadie. Så vi må ikke lade os opsluge af den hype, der kører i højeste gear lige nu. ("AI mangedobler produktiviteten på rekordtid").

Der flourerer mange udokumenterede påstande og vilde overdrivelser.

Kun excellent ledelse (incl. teknologiledelse, herunder AI) kan medføre sådanne resultater.

Med dette paper vil jeg vise, hvordan ChatGPT ikke lever op til hypen, øger risiko, og ikke leverer på innovation, elementær cybersecurity, risikobegrænsning, Privacy compliance og Operational Excellence.

ChatGpt kan meget vel reducere produktiviteten, medmindre det implementeres og anvendes rigtigt.

Er målet virkelig "Mere og hurtigere"? Den nuværende generation af AI kan levere skalering, altså udføre opgaver hurtigere, under inddragelse af mange parametre, og flere iterationer. Det er altså stadig bare IT som vi kender det. Jeg viser her, hvordan ensidigt fokus på skalering og overdrevne forventninger kan være kontraproduktivt.

Lad os se på toppen af hypen, som alle forventes at bøje sig for. ChatGpt er et godt eksempel. Jeg har spurgt ChatGpt om forskellige vitale emner inden for Privacy (GDPR) og informationssikkerhed (ISO 27001) og viser i dette paper, hvordan ChatGpt ikke er compliant til Privacy lovgivningen i GDPR, og bryder elementær informationssikkerhed, til tider leverer misinformation, og dræber innovation.

– altså strider ChatGpt imod Operational Excellence. Det kan føre til kritiske fejltrin, og reducere produktiviteten.

For nu at få de hele ned på jorden, må vi være realistiske om **hvorfor, hvor og hvad AI kan bruges til** i vores organisation, og gå klart efter de innovative forretningsmæssige fordele, der fører videre, samtidig med risikobegrænsning. Her er overvejelserne fra en af de store strategiske tænkere, Bill Gates:

"Vi har formentlig to omdrejninger mere på skalering, hvorved vi får adgang til videodata og bliver meget gode til syntetiske data, som vi sandsynligvis kan skalere op to gange mere. Det er ikke den mest interessante dimension. Den mest interessante dimension er det, jeg kalder metakognition, hvor man forstår, hvordan man tænker på et problem i bred forstand og træder tilbage og siger: "Okay, hvor vigtigt er dette svar? Hvordan kunne jeg tjekke mit svar? Hvilke eksterne værktøjer ville hjælpe mig med dette?"

Jeg mener, det er klart, at AI og store generative modeller som f.eks ChatGPT også på kort sigt har mange muligheder og anvendelser, der understøtter Operational Excellence og innovation og booster forretningsværdi, under fornuftig risikobegrænsning. Det kan lade sig gøre. Der er f.eks. dokumenteret særdeles gode resultater ved beregninger inden for fysik, kemi, økonomi og medicinsk forskning. Også ChatGPT kan købes til On Premise hosting, med eget datagrundlag og egen integration til forretningsssystemer. En konstruktion, der til en vis grad ligner MS Copilot.

Det medfører, at virksomheden selv står for Governance, implementering, drift, risiko, Privacy og sikkerhed.

Dette paper omhandler OpenAI's internetbaserede ChatGPT tilbud.

Afgørende nu og senere er, at vores AI strategi og policies klart fortæller, hvor AI skal sættes ind, hvordan AI løsninger skal understøtte forretningsmål, hvordan der implementeres Risk Management og AI Governance, Data Governance og et AI Management System med controls og målinger, så AI løsningerne dokumenterer klar forretningsværdi under risikobegrænsning.

Det er hvad vi i Symbic kalder BALANCED AI OPERATIONAL EXCELLENCE

1. Privacy – compliance til GDPR

ChatGPT's datagrundlag: Scraping – ”maskinhøst” – af data på internettet

Spørgsmål: Præcis hvilke regler for privacy følger ChatGPT under Scraping af webindhold? ChatGpt svarer:

ChatGPT følger reglerne for beskyttelse af personlige oplysninger og privatlivets fred, når det skraber webindhold. Dette inkluderer at respektere hjemmesidens vilkår og betingelser for brug, undgåelse af at indsamle personlige oplysninger uden samtykke og sikring af fortroligheden af eventuelle data, der indsamles under skrabningsprocessen. ChatGPT overholder også lovgivningen om beskyttelse af personlige oplysninger, herunder GDPR og andre relevante databeskyttelseslove.

Spørgsmålet besvares ikke .ChatGpt er internetbaseret, og kører rundt med en ”skraldevogn” der indsamler informationer om mennesker, og hvad folk nu har skrevet. Det er hverken troværdigt eller lovligt iflg. GDPR. Persondata føjes til ChatGpt's datagrundlag, uden at have afgivet tilladelse hertil).



Der er en elefant i rummet

Den skal væk...

Explainer: GPT er forkortelse for Generative Pretrained Transformation.

Det Europæiske Databeskyttelsesråd (EDPB)

har 24. maj 2024 offentliggjort sin [ChatGPT Taskforce Report](#), og **der er en stor elefant i rummet**. Det kan forklares således:

(ChatGPT er udviklet af virksomheden OpenAI, og anvendes out-of-the-box som en internet-baseret app)



Legitim interesse iflg. GDPR

Med hensyn til webskrabning og "indsamling af træningsdata, forbehandling af data og træning" noterer rapporten, at OpenAI er afhængig af *legitim interesse* i at indsamle og behandle personoplysninger for at træne ChatGPT.. Om det står der i rapporten:

1. *"Det skal erindres, at den retlige vurdering af artikel 6, stk. 1, litra f), GDPR bør baseres på følgende kriterier:*

i) eksistensen af en legitim interesse,

ii) nødvendigheden af behandling, da personoplysningerne skal være tilstrækkelige, relevante og begrænset til, hvad der er nødvendigt i forhold til de formål, hvortil de behandles, og iii) interesseafvejning (...)" Og også

2. *"Som allerede anført af den tidligere Artikel 29-Gruppe (vedr. Retningslinjer for gennemsigtighed i henhold til GDPR), spiller tilstrækkelige garantier en særlig rolle med hensyn til at mindske unødige konsekvenser for registrerede og kan derfor ændre afvejningstesten til fordel for den registeransvarlige. Selv om vurderingen af lovligheden stadig er genstand for igangværende undersøgelser, kan sådanne garantier bl.a. være tekniske foranstaltninger, der definerer præcise indsamlingskriterier og sikrer, at visse datakategorier ikke indsamles, eller at visse kilder (såsom offentlige profiler på sociale medier) udelukkes fra dataindsamling. Desuden bør der træffes foranstaltninger til at slette eller anonymisere personoplysninger, der er indsamlet via webscraping inden uddannelsesfasen."*

Hvad der kort sagt siges er, at skrabning for at træne ChatGPT under påberåbelse af legitim interesse kan være mulig, hvis der anvendes tekniske foranstaltninger, som dem, der er beskrevet ovenfor.

Problemer:

Hvis OpenAI påberåber sig legitime interesser, skal gennemsigtighedsforpligtelserne overholdes

⇒ Artikel 14.5 (b) i GDPR siger, at i tilfælde, hvor det ikke er muligt at underrette registrerede om de oplysninger, der behandles (såsom i forbindelse med scraping), "skal den dataansvarlige træffe passende foranstaltninger for at beskytte den registreredes rettigheder og friheder og legitime interesser, herunder at gøre oplysningerne offentligt tilgængelige."

⇒ **ChatGPTs data er ikke offentligt tilgængelige!**

⇒ **I modsætning til en søgemaskine kan den registrerede ikke udøve simple registreredes rettigheder, for eksempel retten til at blive glemt.**

Personlige data anonymiseres ikke, da systemet stadig udsender oplysninger om mennesker.

Der indgås forskellige licensaftaler med websteder som Reddit og andre platforme, der indeholder personlige data.

Det er således klart, at OpenAI ikke undgår personlige data eller planlægger at undgå det.

ChatGpt er uanvendelig til forretningsformål. Byg en GPT løsning inhouse

ChatGpt anvender brugernes informationer ved at føje dem til modellen. Derved er persondata og firmaets forretningshemmeligheder ikke længere hemmelige. Vi ved ikke hvad der videre sker med data. Derfor: Byg selv en chatbot løsning. Der er flere tilbud i markedet på LLM platforme, der evt. kan udvikles en egen løsning på on premise, som tilbyder sikkerhed og GDPR compliance.

2. Informationssikkerhed – compliance til ISO 27001

Jeg har undersøgt nogle sikkerhedsproblemer med ChatGpt, og giver ChatGPT lejlighed til at svare på spørgsmål om egen sikkerhed, baseret på mine iagttagelser.

Explainer: Hvad gør informationssikkerhed?

Informationssikkerhed beskytter primært

Konfidentialitet – Integritet – Adgang

Informationsintegritet, definition (ISO 27000)

Der skal vises evidens for opfyldelse af hvert af disse tre kriterier:

- Information er ikke modificeret under lagring eller i transit
- Der udføres kun autoriserede modifikationer
- Data er akkurate, autentiske, og sikret imod uautoriseret adgang

Spørgsmål: Er ChatGPT compliant med de grundlæggende regler for informationssikkerhed, særligt kravet om informationers integritet, som krævet i sikkerhedsstandard ISO/IEC 27001, klausul 6.1.2?

Ja, ChatGPT er designet med sikkerhed i tankerne og overholder de grundlæggende regler for informationssikkerhed, herunder kravet om at opretholde integriteten af informationer, som angivet i ISO/IEC 27001 klausul 6.1. Systemet er designet til at beskytte dataene og sikre, at informationerne forbliver uændrede og pålidelige under hele kommunikationsprocessen. Der er også implementeret sikkerhedsforanstaltninger for at forhindre uautoriseret adgang eller ændring af dataene.

Spørgsmål: Hvordan kan sikkerhedsprofessionelle verificere integritet en af informationer, der er oplyst af ChatGPT?

Sikkerhedsprofessionelle kan undersøge integriteten af oplysningerne, der er oplyst af ChatGPT, ved at anvende følgende metoder:

1. Verifikation af kilden: Det er vigtigt at undersøge, hvor oplysningerne stammer fra, og om kilden er pålidelig og troværdig. Hvis oplysningerne kommer fra en ukendt eller tvivlsom kilde, bør de behandles med forsigtighed.
2. Faktatjekning: Sikkerhedsprofessionelle kan foretage faktatjek for at verificere de oplysninger, der er blevet givet af ChatGPT. Dette kan omfatte at søge efter pålidelige kilder og dokumentation for at bekræfte sandheden i oplysningerne.
3. Kontekstualisering: Det er vigtigt at undersøge den kontekst, hvori oplysningerne blev givet af ChatGPT. Nogle gange kan kunsten at nuancere og forstå konteksten af en samtale hjælpe med at afgøre, om oplysningerne er korrekte eller ej.
4. Sammenligning med andre kilder: Sikkerhedsprofessionelle kan sammenligne oplysningerne givet af ChatGPT med andre pålidelige kilder for at se, om der er konsistens i informationen. Hvis der er uoverensstemmelser mellem oplysningerne, kan det være et tegn på, at informationen ikke er pålidelig.
5. Brug af teknologiske værktøjer: Der findes også teknologiske værktøjer, der kan hjælpe med at verificere oplysningerne, der er blevet givet af ChatGPT, såsom tekstanalyseværktøjer og machine learning-algoritmer, der kan identificere falske oplysninger.

Ved at anvende disse metoder kan sikkerhedsprofessionelle bedre vurdere integriteten af oplysningerne, der er blevet oplyst af ChatGPT, og træffe informerede beslutninger baseret på troværdige og pålidelige oplysninger.

(Her går spørgsmålet specifikt efter verifikation, der kræver evidens.) Men svaret her er i kategorien "varm luft".

- 1) *Verifikation af kilden er noget, som enhver kan gøre på kort tid, såfremt kilden er angivet. Det er et must. Men kilder angives ikke i ChatGPT, så verifikation er ofte ikke mulig. Det samme gælder dermed også pkt. 4.*
- 2) *Faktacheckning af ChatGPT uden kilder er temmelig ressourcekrævende, og der savnes et framework hertil.*
- 3) *Svaret er ude af scope. Dataintegritet er ethovedkrav i ISO 27001, der strækker sig ud over mange controls.*
- 4) *Brugeren kan ikke sammenligne kilder, idet ChatGPT ikke kan oplyse kilder.*
- 5) *At undersøge ét AI system med et andet AI system svarer til at undersøge én black box med en anden black box. Verifikatoren er sat uden for kontrol, og kan ikke finde evidens.*

Spørgsmål: Hvordan er oplysninger, der gives af ChatGPT beskyttet imod misinformation?

ChatGPT beskytter oplysninger mod misinformation ved at trække på en række forskellige kilder og ressourcer for at sikre, at de oplysninger, der gives, er nøjagtige og pålidelige. ChatGPT bruger avancerede algoritmer til at analysere og vurdere informationen, før den præsenteres for brugeren. Derudover kan brugere også bidrage til at korrigere eventuelle fejl eller misforståelser i oplysningerne, hvilket hjælper med at forbedre kvaliteten og nøjagtigheden af de oplysninger, der gives. Endelig kan brugere også få adgang til yderligere kilder og ressourcer for at verificere oplysningerne, hvis de er i tvivl om deres nøjagtighed. På denne måde arbejder ChatGPT aktivt på at beskytte brugere mod misinformation ved at sikre, at de oplysninger, der gives, er pålidelige og korrekte.

Ifølge dette svar skal man altså blot stole på at ChatGPT oplysninger er valide. Ifølge ovenstående er det en trossag. Vi må konkludere, at ChatGpt's svar kan være misinformation

Spørgsmål i forlængelse af ovenstående: Men ChatGPT kan ikke angive kilder til informationer?

Ja, det er korrekt. ChatGPT er en AI-model, der genererer svar baseret på den information, den er blevet trænet på, men den kan ikke angive specifikke kilder eller referencer til de oplysninger, den giver. Det er vigtigt at huske på, at ChatGPT er et værktøj til generel information og ikke bør bruges som en primær kilde til videnskabelige eller akademiske oplysninger. Det er altid en god idé at dobbelttjekke oplysninger fra flere kilder, især når det drejer sig om vigtige eller komplekse emner.

Ja, informationer er hurtigt fremme med ChatGPT, også misinformation

Out-of-the-box synes det ikke attraktivt at implementere et hurtigt informationssystem, der bedst kan sammenlignes med at køre 250 km/t i en bil med defekt styretøj og bremses. – (Bla. Novo Nordisk har lavet sin egen interne GPT). Det er ikke uoverkommeligt, og der findes også færdige inhouse-løsninger i markedet.



UPS - det gik da ellers hurtigt

Med hovedet under armen: Hvem ville køre 250 km/t i en bil med defekt styretøj og bremses?

De uheldige helte: To New York advokater, der gik i retten med ChatGPT og hovedet under armen.

Ikke nok med at de tabte deres sag, nej de blev også idømt klækkelige bøder for at anvende ChatGPT under udarbejdning af sagsfremstilling, der henviste til tidligere domme i sager, der var frit opfundet af ChatGPT.

Det har medført store tab for det advokatfirma, der har udstyret medarbejderne med ChatGPT misinformation.

(Reuters)

“New York lawyers sanctioned for using fake ChatGPT cases in legal brief”

Flere problemer med datagrundlaget – ”Garbage in- garbage out”

Bias: ChatGPT træningsdata er påvirket af holdninger og normer hos de der træner datagrundlaget.

Infusion: Da datagrundlaget er internetbaseret, kan data inficeres af f.eks hackere, udenfor vores kontrol.

Ny rapport om AI-sikkerhed



Den internationale videnskabelige (foreløbige) [rapport om sikkerheden ved avanceret AI](#) er udgivet. Den indeholder indsigt fra 75 AI-eksperter, herunder et internationalt rådgivende ekspertpanel udpeget af 30 lande, EU og FN. Den blev offentliggjort for at informere drøftelserne i forbindelse med AI Seoul-topmødet 2024. Nedenfor er nogle højdepunkter i resuméet:

⇒ "Hvis den styres korrekt, kan AI til generelle formål anvendes til at fremme offentlighedens interesse, hvilket potentielt kan føre til forbedret trivsel, mere velstand og nye videnskabelige opdagelser. Funktionsfejl eller ondsindet anvendt kunstig intelligens til generelle formål kan imidlertid også forårsage skade, f.eks. gennem partiske beslutninger i high-stakes-miljøer eller gennem svindel, falske medier eller krænkelser af privatlivets fred.

⇒ Efterhånden som den generelle AI-kapacitet fortsætter med at udvikle sig, kan der opstå risici såsom omfattende virkninger på arbejdsmarkedet, AI-aktiveret hacking eller biologiske angreb, og samfundet mister kontrollen over AI til generelle formål, selv om sandsynligheden for disse scenarier diskuteres blandt forskere. Forskellige synspunkter om disse risici stammer ofte fra forskellige forventninger til de skridt, samfundet vil tage for at begrænse dem, effektiviteten af disse skridt, og hvor hurtigt AI-kapaciteter til generelle formål vil blive avanceret.

⇒ Der er forskellige tekniske metoder til at vurdere og reducere risici fra AI til generelle formål, som udviklere kan anvende, og regulatorer kan kræve, men de har alle begrænsninger. For eksempel er de nuværende teknikker til at forklare, hvorfor AI-modeller til generelle formål producerer et givet output, stærkt begrænsede.

⇒ Fremtiden for AI-teknologi til generelle formål er usikker, idet en lang række forløb synes mulige selv i den nærmeste fremtid, herunder både meget positive og meget negative resultater. Men intet om fremtiden for AI er uundgåeligt. Det vil være samfundets og regeringernes beslutninger, der vil afgøre AI's fremtid. Denne interimrapport har til formål at fremme konstruktive drøftelser om disse beslutninger."

3. ChatGPT dræber innovation og Operational Excellence

En chatbot af ChatGPT typen leverer hurtige middelmådige resultater, der kan fremme ”mental dovenskab” hos brugerne. (Eller utilfredshed.)

Medarbejderne mister ansporingen til at finde på bedre måder at gøre arbejdet, når ledelsen signalerer, at misinformation fra internettet er godt nok for os. Det er meget farligt for virksomhedens produktivtetsudvikling og fremtidige konkurrencekraft. De dygtige og engagerede ledere og medarbejdere søger væk til andre, mere innovative og ansvarlige virksomheder. (Mennesker har en indbygget trang til at stræbe efter noget bedre.)

Det får mig til at tænke på dengang i 1970’erne, da vi blev tilbudt at købe billig -blandet, presset og pumpet - underlødigt kød fra EU’s overskudslagre i kæmpestore konservesdåser. Det udløste opfindelsen af betegnelsen ”Spam” i *Monty Pythons Flying Circus’* satire ”Spam-Spam-Spam”. Vi ønsker vel noget bedre for vore AI løsninger.

Innovation og excellent menneskelig udvikling er afgørende for succes

Læs bogen ”Heartcore” af Mads Øvlisen, mangeårig CEO og bestyrelsesformand for Novo Nordisk. Øvlisen fortæller om ledelsens innovationspolitikker, der har skabt Europas mest værdifulde selskab.

Bogen fås på engelsk på Amazon.com, redigeret af Siegfried W Andersen. Her kan vi lære om ledelse af højeste karat. Og så kan vi tænke over, om ChatGPT middelmådigheden passer ind her.

<https://www.amazon.com/Heartcore-Thoughts-Leadership-Mads-%C3%98vlisen/dp/B0CGYQ2VJG>

udgivet af KF Andersen Leadership Academy

Konklusion

ChatGPT kategoriseres High Risk iflg. EU AI Act, og er uanvendelig til professionel brug

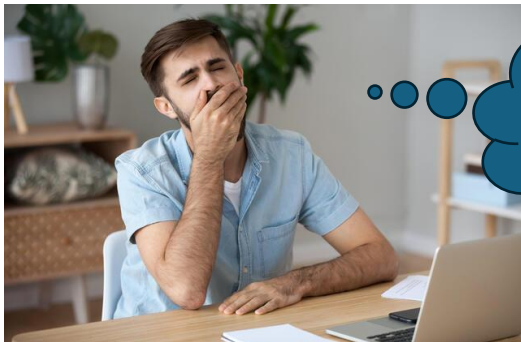
ChatGPT Privacy beskyttelse iflg. GDPR og sikkerhed iflg. ISO 27001 er uacceptabel

Sikkerhedsprofessionelle kan almindeligvis ikke verificere integriteten af informationer, der er oplyst af ChatGpt. Der er altså høj risiko for misinformation. Dertil kommer bias og truslen mod beskyttelse af egne forretningshemmeligheder på nettet

ChatGpt indebærer en trussel imod innovation og engageret virksomhedskultur

Det skal gøres bedre og ansvarligt med BALANCED AI OPERATIONAL EXCELLENCE

Min anbefaling med hensyn til GPT og lignende skal være at tænke i retning af egne inhouse GPT løsninger der opfylder krav til forretningsmæssig og faglig relevans, understøtter og beskytter opnåelse af forretningsmål, og Innovation, risikominimering, informationssikkerhed og compliance til GDPR- og EU AI Act



Det her må da kunne gøres på en bedre måde...

Lær hvordan på vore AI kurser
https://symbic.dk/?page_id=623

Gør noget ved det



Symbic hjælper med AI Operational Excellence

AI og AI GPT policies og governance – AI Management System implementering
Risikovurdering – AI Audit og System Impact Analyse – Compliance til EU AI Act

symbic

Information Governance – Risk Management – AI Management – Information Security og Cybersecurity
Symbic A/S Guldager 19, 3630 Jægerspris – www.symbic.dk – tlf. (AI ansvarlig, CEO Erik J. Andersen) 2344 2040 – email eja@symbic.dk