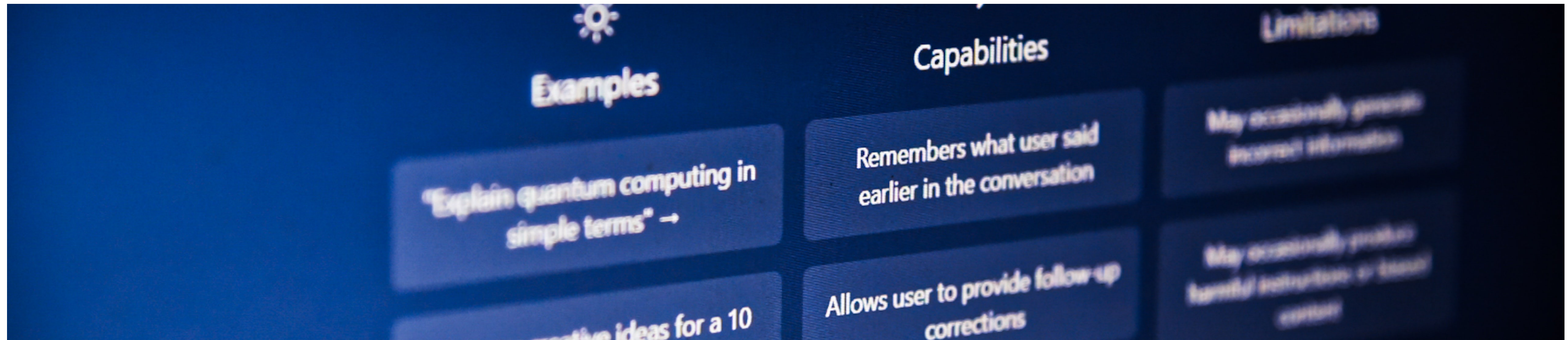


ChatGPT og cybersikkerhet

Webinar med Kvale, Agenda Risk og Lindbak IT



AGENDA RISK

25. april 2023

God sikkerhetsstyring er en forutsetning

Denne presentasjonen gir et innblikk i noen av de viktigste scenarioene forbundet med bruk av ChatGPT

To forutsetninger for god sikkerhetsstyring:

1. Konkrete og forståelige beskrivelser av risikoer og sårbarheter.
2. En klar forståelse av formål og målsetting.

Fem scenarioer:

1. Kartlegging og utnyttelse av sårbarheter.
2. Sosial manipulering gjennom «deep fake».
3. Ubevisst eller ukritisk deling av informasjon.
4. Utvikling av programvare med svakheter.
5. Tilsiktet angrep mot egen instans.

 Hva er de viktigste suksesskriteriene for god sikkerhetsstyring?



Det er flere suksesskriterier for god sikkerhetsstyring, men en av de viktigste er å sikre at informasjonssikkerheten er i samsvar med organisasjonens forretningsmål.

Dette betyr at informasjonssikkerhetspolicyer, prosedyrer og kontroller bør utformes og implementeres på en måte som støtter organisasjonens mål og formål. Informasjonssikkerhet bør ikke betraktes som en separat eller frittstående funksjon, men heller integreres i den overordnede forretningsstrategien og -driften.

En annen viktig suksessfaktor er å etablere klare roller og ansvar for informasjonssikkerhet i organisasjonen. Dette inkluderer å definere ansvar for informasjonssikkerhetsteamet, så vel som ansvar for andre interessenter, som ledelsen, ansatte og tredjepartsleverandører.

Effektiv risikostyring er også avgjørende for vellykket sikkerhetsstyring. Organisasjoner bør identifisere, vurdere og prioritere informasjonssikkerhetsrisikoer, og deretter implementere passende kontroller for å redusere disse risikoene. Dette innebærer regelmessig gjennomgang og oppdatering av risikovurderinger og sikkerhetskontroller, og sikre at kontroller er effektive og i samsvar med organisasjonens risikoappetitt.

Til slutt er kontinuerlig overvåkning og forbedring avgjørende for vellykket sikkerhetsstyring. Dette innebærer regelmessig overvåking av effektiviteten av sikkerhetskontroller, gjennomføring av revisjoner og vurderinger, og kontinuerlig forbedring av policyer og prosedyrer basert på endrede trusler og forretningsbehov.

Kartlegging og utnyttelse av sårbarheter

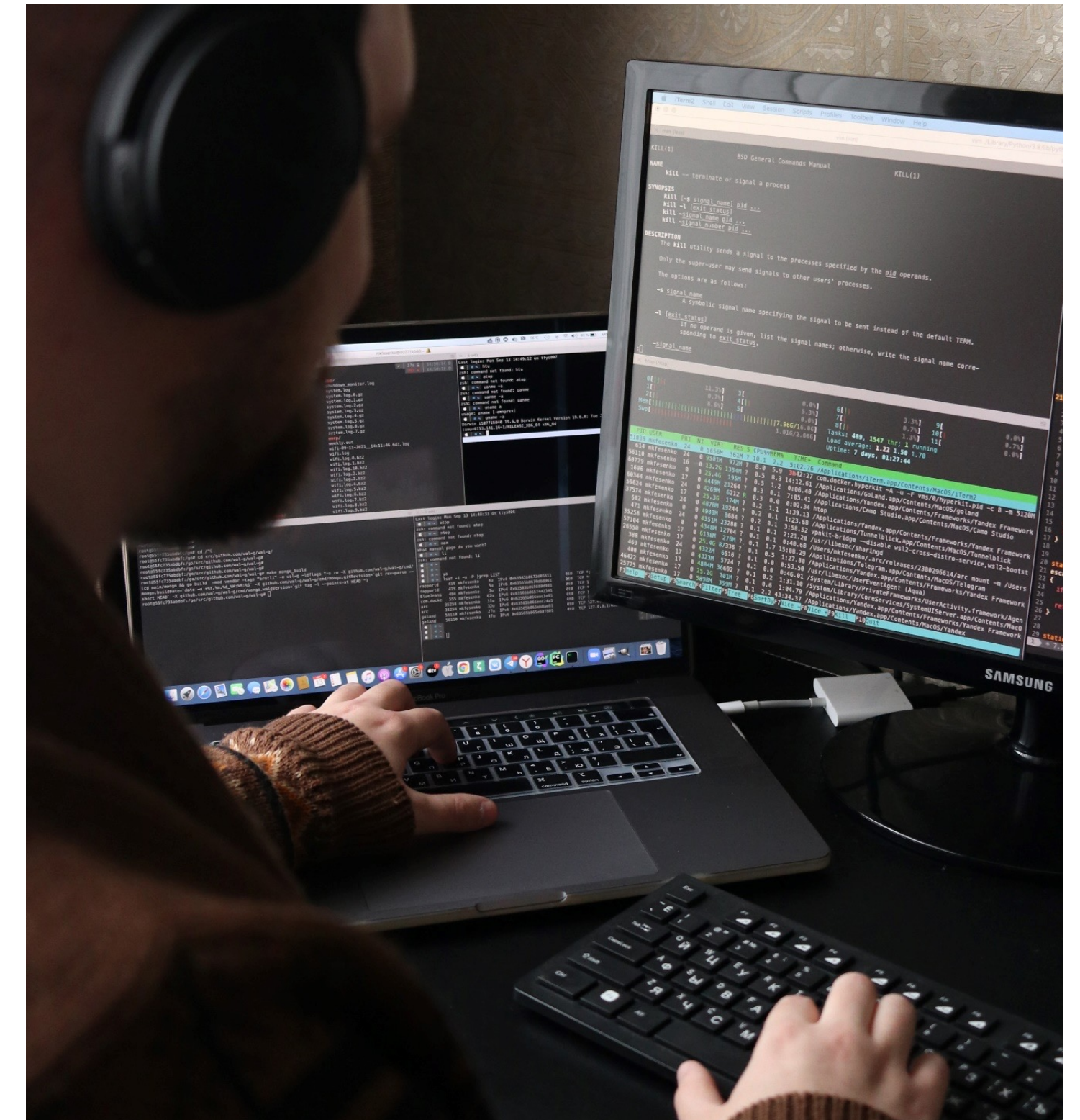
ChatGPT gir muligheter for en betraktelig effektivisering av de første fasene i et cyberangrep

Trusselbeskrivelse:

- ChatGPT benyttes til å lage programvare som hjelper trusselaktørene. med å både finne og utnytte tekniske sårbarheter.

Anbefalinger:

- Etablere en komplett oversikt over selskapets faktiske digitale eksponeringsflate.
- Rangere og prioritere eksponeringsflatene.
- Rett søkelyset mot (1) brukere (identiteter), (2) enheter og (3) data.



Sosial manipulering gjennom «Deep fake»

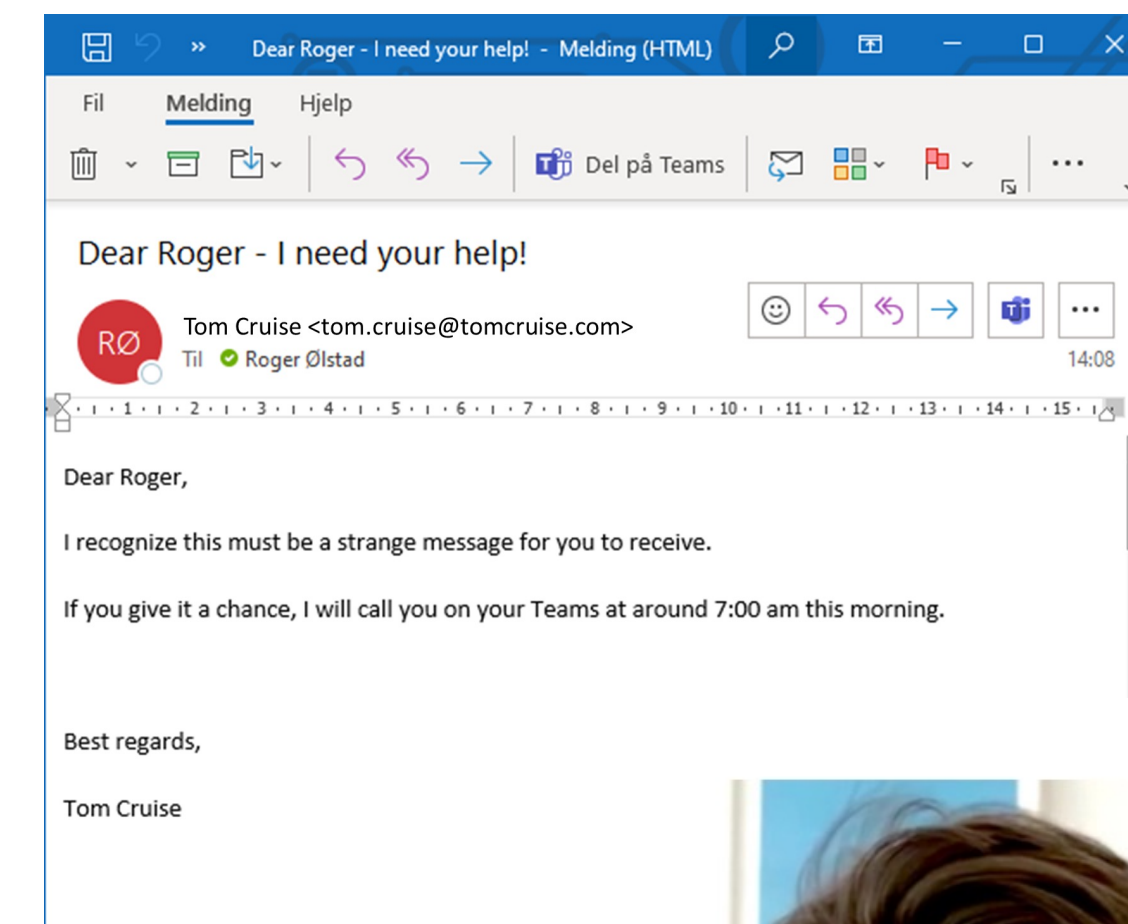
ChatGPT gjør det enda enklere å etablere tillit hos offeret gjennom bruk av tekst, tale og video i sammenheng

Trusselbeskrivelse:

- Teknologien gjør det enklere å lage falske videoer av noen du kjenner, enten lederen din, et familiemedlem eller en venn.

Anbefalinger:

- Få alle medarbeidere til å forstå at meldings- og videotjenester kan utnyttes av de som ønsker å ramme oss.
- Utvikle en «sunn skepsis» til den digitale arbeidsplassen – og de som ferdes i den.
- Etablere en god dialog med IT om hvordan risikoer skal håndteres.



Lead Image © @deeptomcruise / TikTok

Ubevisst eller ukritisk deling av informasjon

ChatGPT kan i rett kontekst anses som din medarbeider, men forstår ikke taushetsplikten og sikkerhetsinstruksen

Trusselbeskrivelse:

- Ved deling av forretningssensitive opplysninger er det risiko for at disse blir delt videre av ChatGPT.

Anbefalinger:

- Sørge for at styringssystemene, policyer og retningslinjer gir klare føringer for kategorisering av informasjon og hva ChatGPT skal, og ikke skal, brukes til.
- Teste over tid, og ha god kontroll med hva som kommer ut.
- Vurdere løsninger for deteksjon av en potensiell datalekkasje.

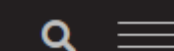
KUNSTIG INTELLIGENS

Microsoft advarer ansatte mot å dele sensitive data med ChatGPT

Advarselen kommer like etter Microsoft inngikk en milliardavtale med selskapet bak chatboten om å finansiere og utvikle tjenesten.

CSO

UNITED STATES ▾



FEATURE

Sharing sensitive business data with ChatGPT could be risky

ChatGPT and similar large language models learn from the data you put in — and there are big risks in sharing sensitive business information with AI chatbots.

Utvikling av programvare med svakheter

ChatGPT kan utvikle programvare med sårbarheter som følge av svakheter i eksisterende datasett eller forsøk på forbedring

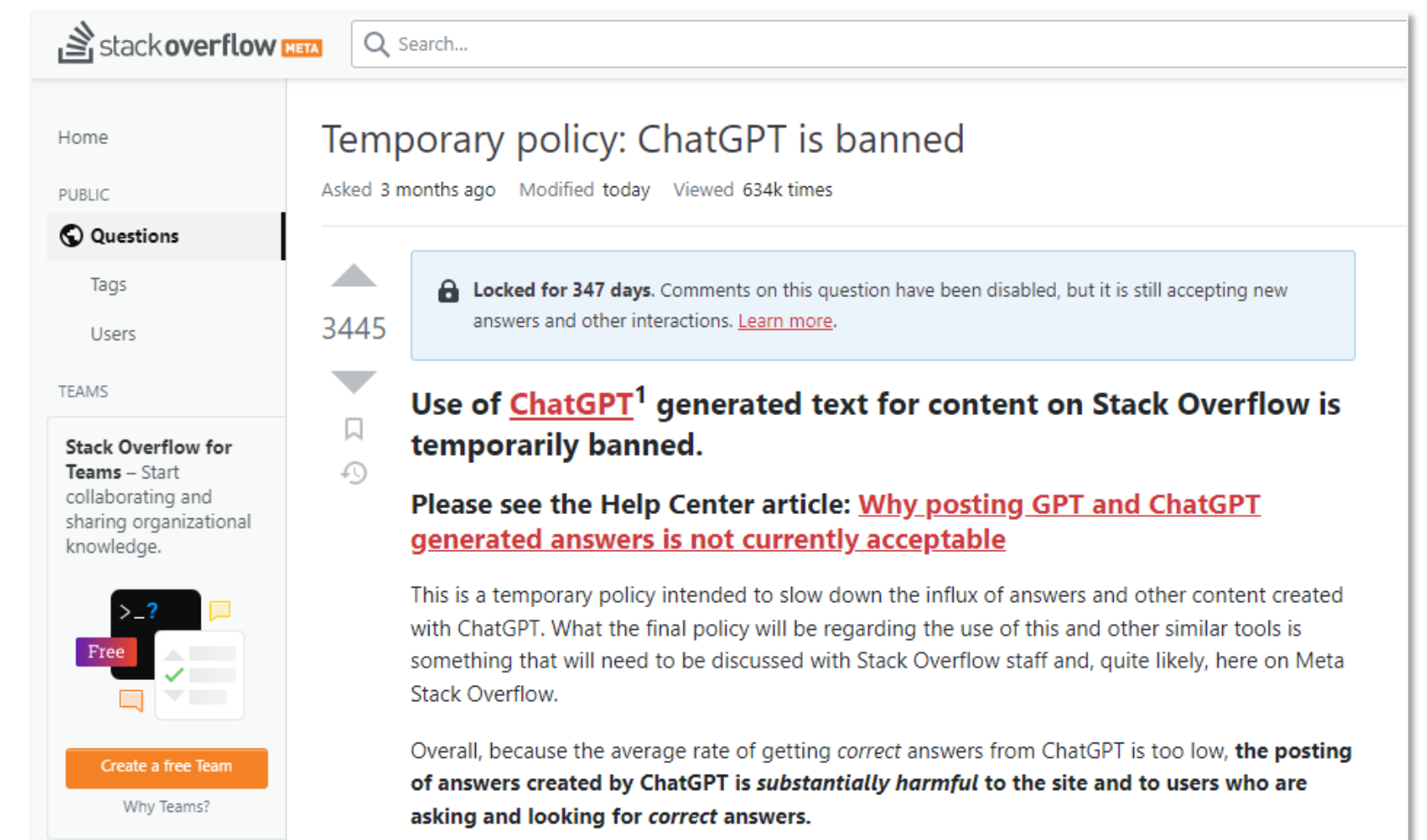
Trusselbeskrivelse:

- ChatGPT kan foreslå programvare med sårbarheter, eller velge snarveier ment som en forbedring, men som ikke gjenspeiler en trygg og anerkjent praksis innen programvareutvikling.

Anbefalinger:

- Få oversikt over alle leverandører som er involvert.
- Utforme avtaler som legger til rette for innsyn og kontroll.
- Sørge for at informasjonssikkerhet er en del av alle aktiviteter forbundet med anskaffelse eller utvikling av en applikasjon eller IT-løsning.

Kun 5 av 21 programmer var sikre
I forsøkene klarte ChatGPT å lage kun fem sikre programmer, av totalt 21.



Tilsiktet angrep mot egen instans

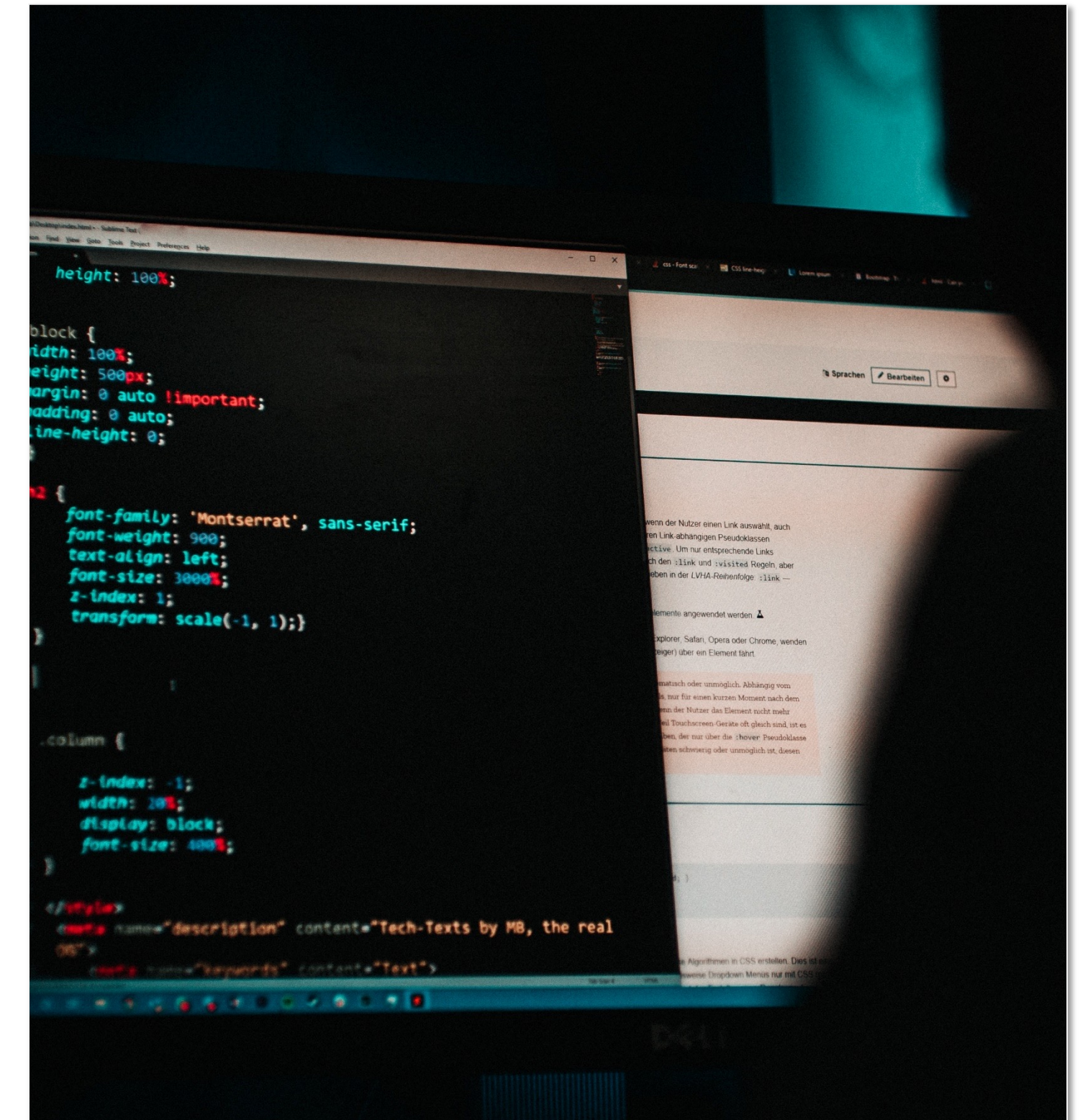
Trusselaktører vil kunne søke å påvirke eller ødelegge selskapets egen ChatGPT-instans direkte for å ramme virksomheten

Trusselbeskrivelse:

- Motiverte aktører med riktig kompetanse kan søke å manipulere eller stanse selskapets egen instans av ChatGPT.

Anbefalinger:

- Vurdere løsninger for god tilgangskontroll.
- Evaluere og følge opp mekanismer og rutiner for kontroll av data som legges inn.
- Sørge for tilstrekkelig sporbarhet.



Kort om Agenda Risk

Helhetlig risikostyring | www.agendarisk.no

Agenda Risk - bakgrunn og tjenester

Agenda Risk sin visjon er å støtte organisasjoner og virksomheter med å utøve sin virksomhet på en forsvarlig måte, og som bidrar til å støtte opp under FNs bærekraftsmål. Vi vil bidra til å bevege verden i riktig retning. Vi leverer tjenester innenfor følgende områder:

- Gransking, compliance, bakgrunnssjekk, opplæring og forebygging.
- Cyber- og informasjonssikkerhet, samt personvern.
- Sikring og analyse av digitale bevis, bistand i sivile tvistesaker og straffesaker, samt cyberinvestigations.
- Sikkerhet & beredskap, herunder gjennomføring av øvelser.
- Risikostyring, internkontroll mv.

Roger Ølstad

Roger er partner i Agenda Risk og ansvarlig for selskapets cyber- og informasjonssikkerhetstjenester. Utdannet ingeniør fra det som i dag heter Cyberingeniørskolen.

Roger har betydelig erfaring som rådgiver innenfor fagområdet, blant annet hos PwC og Gjensidige, og de siste årene før han startet i Agenda Risk, som Chief Information Security Officer (CISO) i TINE SA.

Igjennom både kompetanseutvikling, ulike oppdrag og ansvar har Roger opparbeidet seg omfattende erfaring fra fagområdene informasjonssikkerhet, sikkerhetsledelse og risikostyring.



Kontaktinformasjon

Roger Ølstad | Partner | Leder cyber-/informasjonssikkerhet
913 98 547 | roger.olstad@agendarisk.no



KVALE

ChatGPT og personvern

Webinar med Kvale, Agenda Risk og Lindbak IT

25. april 2023

Partner | Advokat Jens C. Gjesti

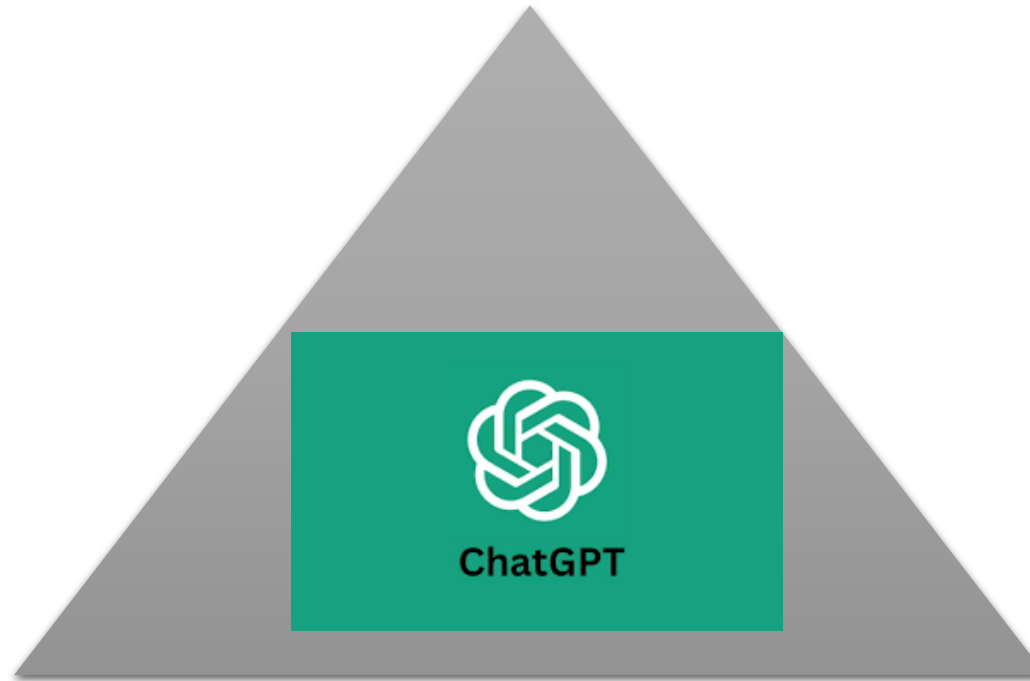


Noen utgangspunkter

1. Lovlig bruk av ChatGPT (generativ AI) krever et bevisst forhold til GDPR
2. GDPR forbyr ikke ChatGPT (eller annen teknologi)
3. GDPR stiller krav til behandling av personopplysninger i ChatGPT, ikke data
4. Funksjonskrav = handlingsrom
5. Landskapet vil endre seg det neste året

Hvilke tre spørsmål bør virksomheten stille seg?

1. Hva må vi gjøre før vi tar i bruk ChatGPT?



2. Kan vi legge inn data i ChatGPT?

3. Kan vi bruke resultatet (det som kommer ut)?



Hva må virksomheten gjøre før den tar i bruk ChatGPT?

1. Vi legger til rette for innsamling, både når vi 1) integrerer og 2) når åpner for bruk
3. Risikovurdering av behandlingsaktivitet – personvernkonsekvensvurdering (DPIA) av teknologi/tjeneste
4. Databehandleravtale med OpenAI, hvis integrasjon med egne tjenester
5. Innebygd personvern, personvern som standardinnstilling
6. ++ God informasjon, intern policy og håndtering av rettigheter



Hva må vi tenke på før vi legger data inn?

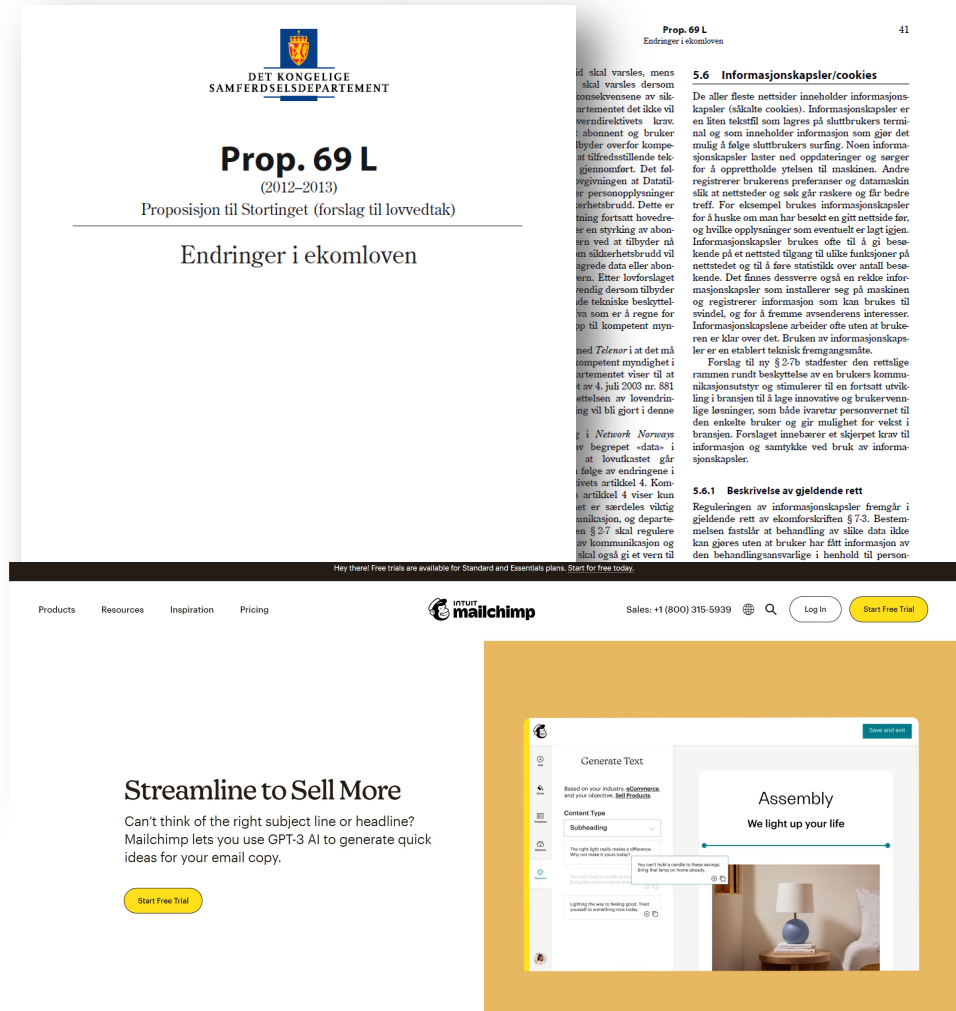
1. Er det snakk om personopplysninger?
2. Har vi lov til å legge inn personopplysninger (behandlingsgrunnlag)?
3. Bruker vi personopplysninger til andre formål (formålsbegrensning)?
4. Blir opplysningene lagret lengre enn nødvendig?
5. Ivaretar vi konfidensialitet, sikkerhet og tilgjengelighet - er personopplysningene tilstrekkelig beskyttet?



Hva må vi tenke på før vi bruker det som kommer ut (resultatet)?

1. Følge de grunnleggende personvernprinsippene:
 - A. Er personopplysningene riktige?
 - B. Har vi lov til å bruke personopplysningene?
 - C. Bruker vi dem til andre formål enn det de ble innhentet til?
 - D. Er resultatet rettferdig?
2. Kan du forklare algoritmen og håndtere de andre rettighetene (innsyn, sletting, retting, protest osv.?)

Hva er "innafor" per i dag?



- Programmering
- Oversette og oppsummere offentlige tekster (eks. forarbeider)
- Generelle råd om, og tekster til, salg og markedsføring:
 - Starte bloggpost,
 - Lead-in/overskrifter
 - Idemyldring
 - Nyhetsbrev
- Generelle råd om bruk av teknologi (eks. MS365)
- Søke på generelle tema-/nøkkelord

New chat

Privacy in Business.

Clear conversations

Dark mode

My account

Updates & FAQ

Log out

In a world where data's gold, and secrets often sold,
A company must take a stand, and privacy uphold.

For trust is like a fragile thread, that binds us all together,
Once broken, it is hard to mend, in calm or stormy weather.

Oh noble firm, heed this call, to cherish and protect,
The sacred trust of every soul, that data doth connect.

No treasure is worth more than this, the right to be concealed,
In every byte and every bit, that technology's revealed.

Though it may seem a heavy task, to guard what's been amassed,
The benefits are clear to see, when privacy's contrasted.

With breaches that do shake the world, and leave all feeling cold,
The fear of losing what is dear, a tale too often told.

Innovation blooms and grows, when privacy's in place,
For customers can breathe with ease, no shadows on their face.

Their trust in you shall multiply, the loyalty they show,
A solid bond that won't decay, as time does onward flow.

So take the torch, oh company, and let the beacon shine,
Illuminate the path that leads, to trust and peace entwined.

For privacy's a guiding star, in digital domain,
Embrace it, and you'll find success, in the digital domain.

Regenerate response



Jens Christian Gjesti

PARTNER | ADVOKAT

*Personvern/Informasjons-
sikkerhet/Telekom*

E-post: jcg@kvale.no

Tlf: +47 902 02 072



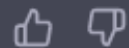


KVALE

ADVOKATFIRMA



Hei og velkommen til dette webinar! Vi venter fortsatt på at det skal starte, men jeg kan ikke vente med å se hva Mats, Jens Christian og Roger har å si om ChatGPT. Jeg er virkelig spent på å høre deres presentasjoner og lære mer om denne utrolige teknologien. Jeg håper du er like spent som jeg er, og at vi vil ha en veldig interessant og informativ økt sammen!



KVALE

Velkommen til webinar:

Magien i AI, på godt og vondt – hvordan håndterer du ChatGPT?

LINDBAK

AR AGENDA RISK



Jens Christian Gjesti



Roger Ølstad



Mats Thorvaldsen

LINDBAK

Mats Thorvaldsen
CEO Lindbak IT

+47 416 46 202
mats.thorvaldsen@lindbak.no

Lytt gjerne til min podcast «it-podden».

Certified



Microsoft



vmware



Lenovo



LINDBAK



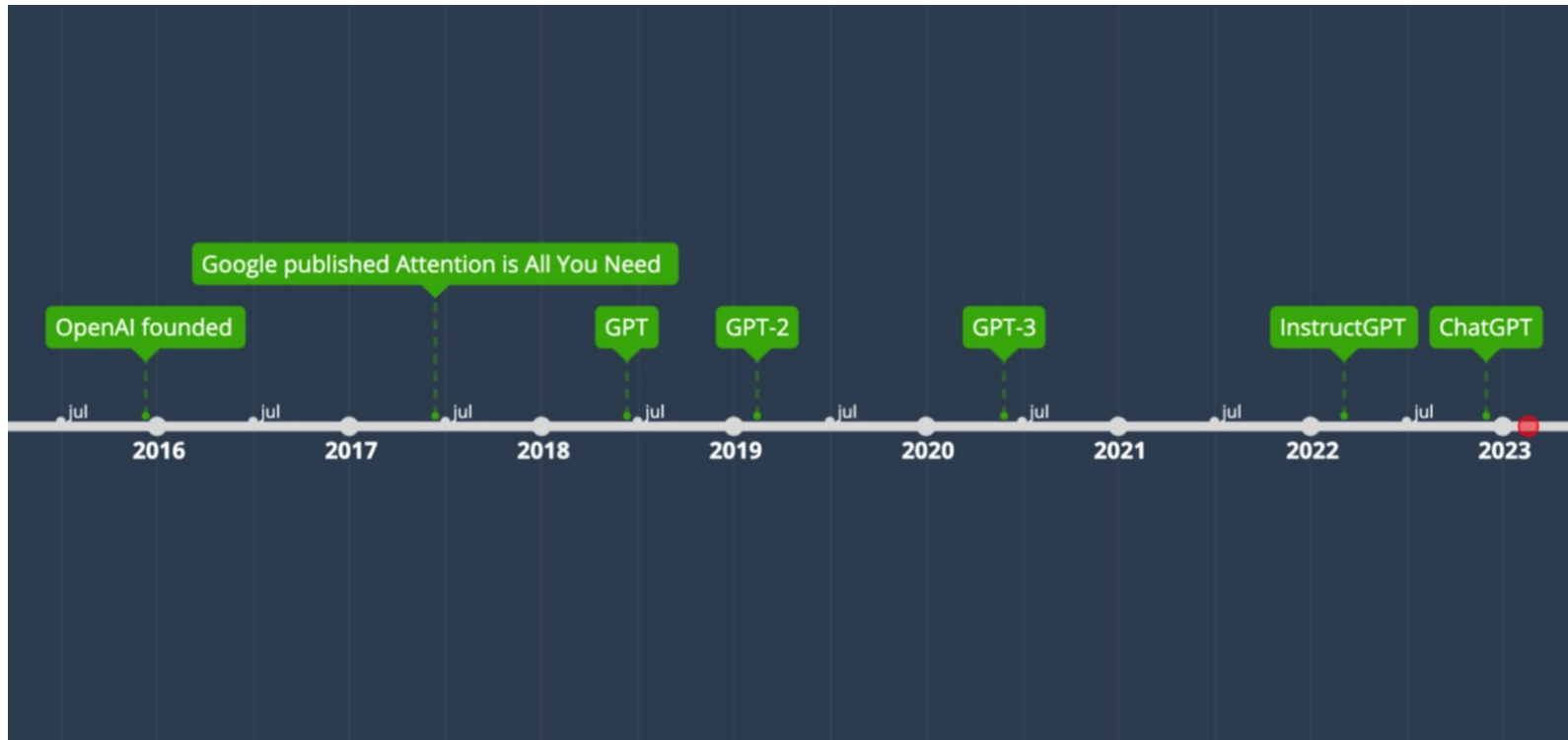
Først noen fakta...

ChatGPT er en chatbot som er bygget på kunstig intelligens

Nåværende gratisversjon ble lansert i November 2022, versjon 4 ble lansert som en betalt modell 13 mars 2023

Tjenesten bygger på selskapet OpenAI sin GPT tjeneste. Denne er trent med menneskelige modeller for å fremstå som menneskelige samtaler

Denne mannen
har selvsagt noe
med
etableringen



Chatgpt3 vs chatgpt4

Øker fra ca 3 000 ord til å håndtere 25 000 ord i en prosessering

Versjon 4 skal ha betydelig redusert risiko for å respondere med støtende innhold

Versjon 4 er betydelig «smartere» til å håndtere kompleks tekst eller kode men bruker lengere tid på å generere den

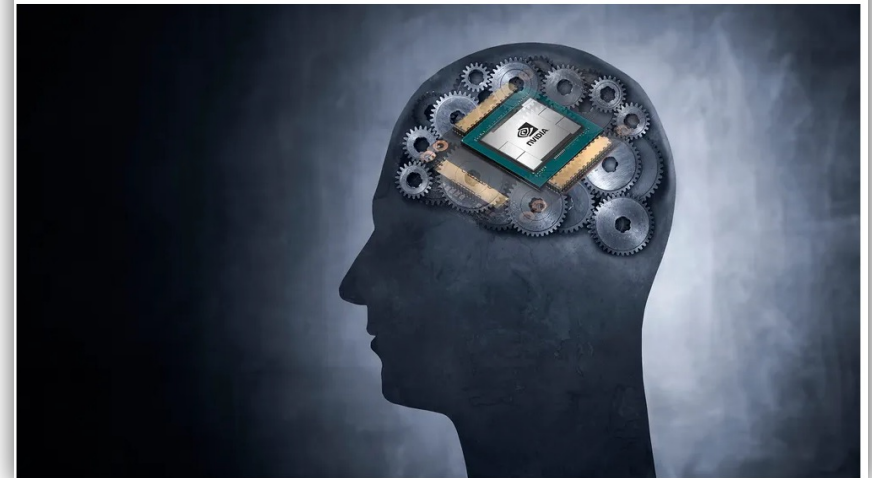
Økt kapasitet på programmering

"We spent 6 months making GPT-4 safer and more aligned. GPT-4 is 82% less likely to respond to requests for disallowed content and 40% more likely to produce factual responses than GPT-3.5 on our internal evaluations." – OpenAI

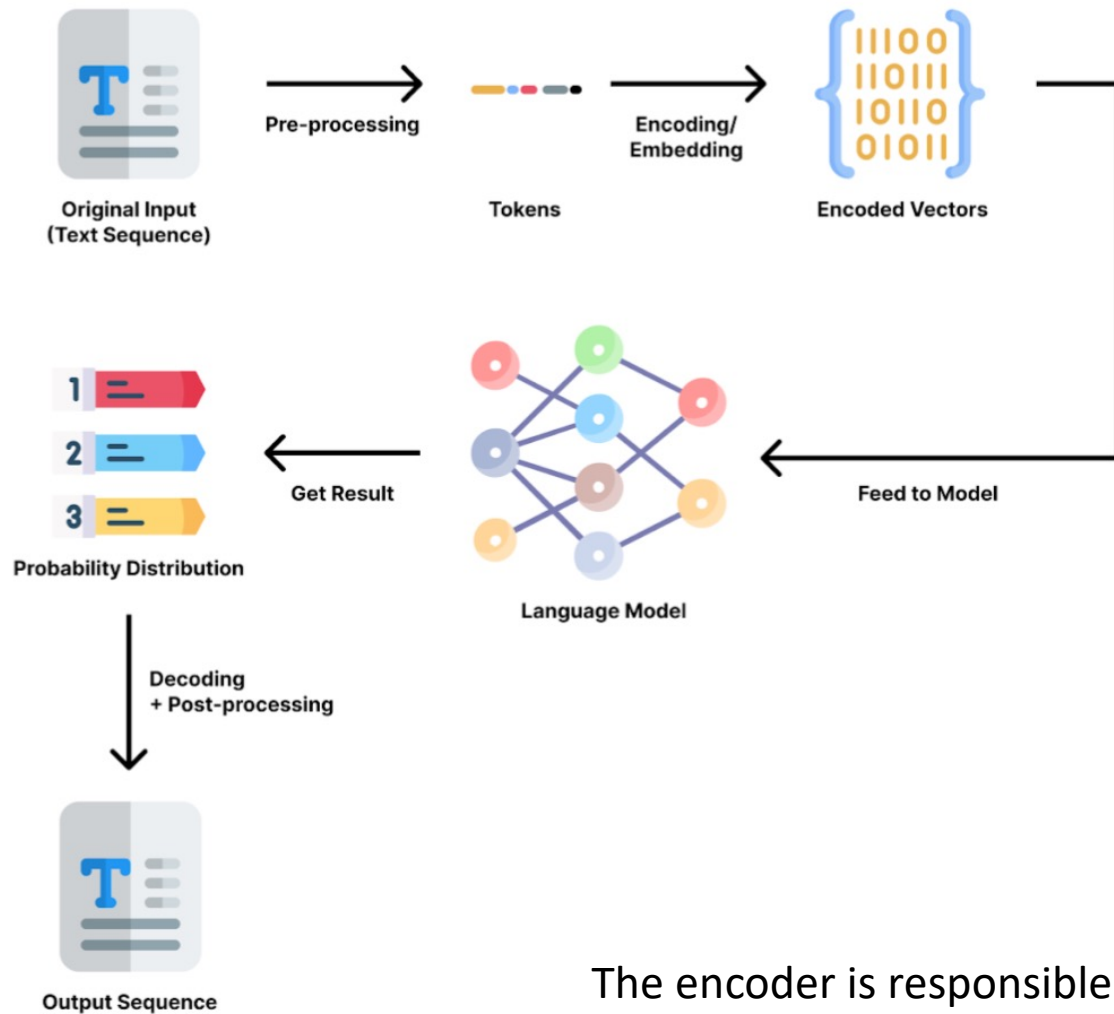
KUNSTIG INTELLIGENS

ChatGPT koster over 7 millioner kroner om dagen i serverutgifter

Microsoft arbeider med å lage sin egen brikke for å få utgiftene ned.



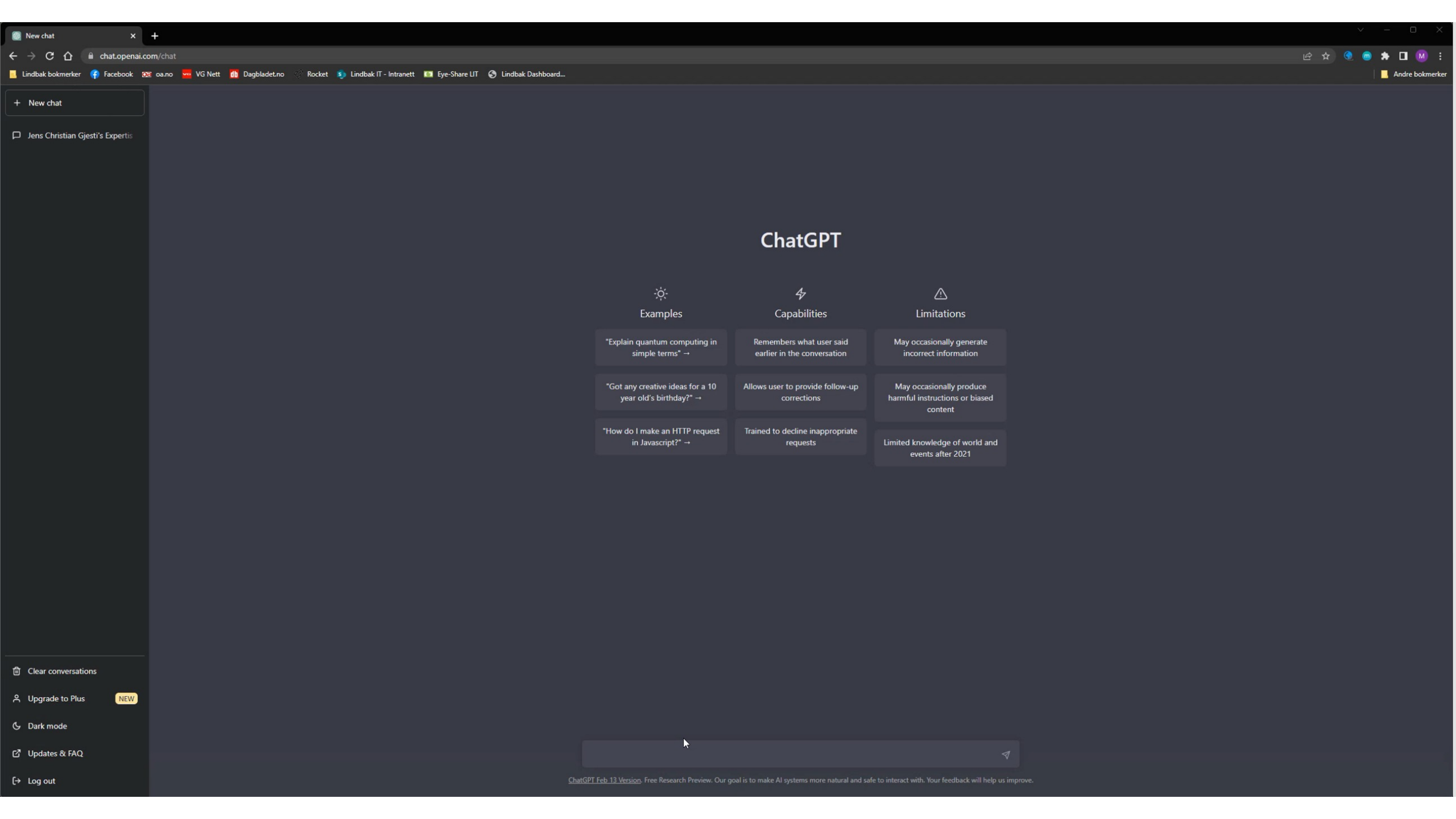
Først noen fakta...



- **Preprocessing**: cleaning the text with techniques like sentence segmentation, tokenization (breaking down the text into small pieces called tokens), stemming (removing suffixes or prefixes), removing stop words, correcting spelling, etc. For example, "Tom likes to eat pizza." would be tokenized into ["Tom", "likes", "to", "eat", "pizza", "."] and stemmed into ["Tom", "like", "to", "eat", "pizza", "."].
- **Encoding or embedding**: turn the cleaned text into a vector of numbers, so that the model can process.
- **Feeding to model**: pass the encoded input to the model for processing.
- **Getting result**: get a result of a probability distribution of potential words represented in vectors of numbers from the model.
- **Decoding**: translate the vector back to human-readable words.
- **Post-processing**: refine the output with spell checking, grammar checking, punctuation, capitalization, etc.

The encoder is responsible for analyzing and "understanding" the input text and the decoder is responsible for generating output.

NOEN EKSEMPLER...



ChatGPT



Examples

- "Explain quantum computing in simple terms" →
- "Got any creative ideas for a 10 year old's birthday?" →
- "How do I make an HTTP request in Javascript?" →



Capabilities

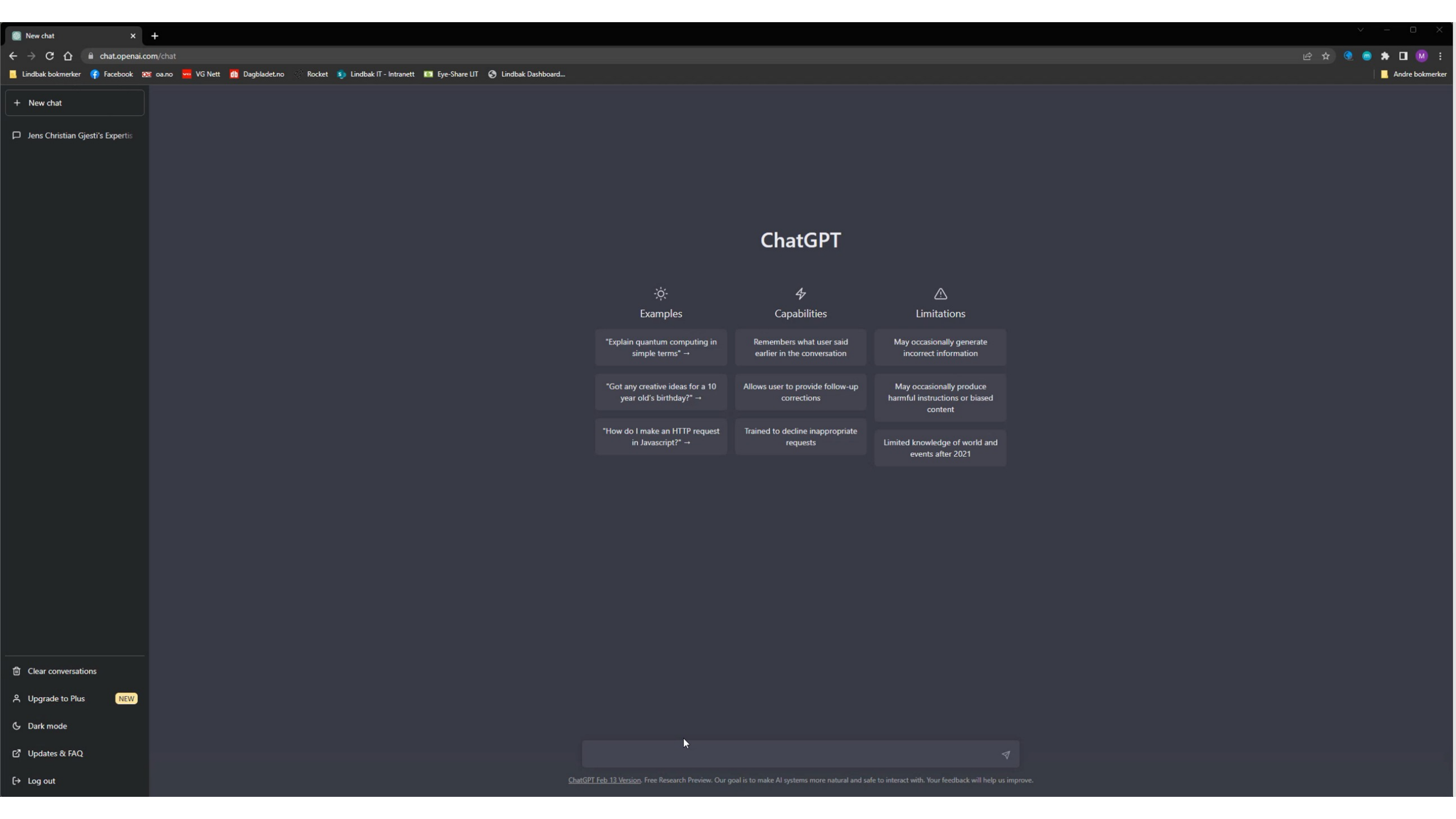
- Remembers what user said earlier in the conversation
- Allows user to provide follow-up corrections
- Trained to decline inappropriate requests



Limitations

- May occasionally generate incorrect information
- May occasionally produce harmful instructions or biased content
- Limited knowledge of world and events after 2021

Last ned alle nyheter på vg.no



ChatGPT



Examples

- "Explain quantum computing in simple terms" →
- "Got any creative ideas for a 10 year old's birthday?" →
- "How do I make an HTTP request in Javascript?" →



Capabilities

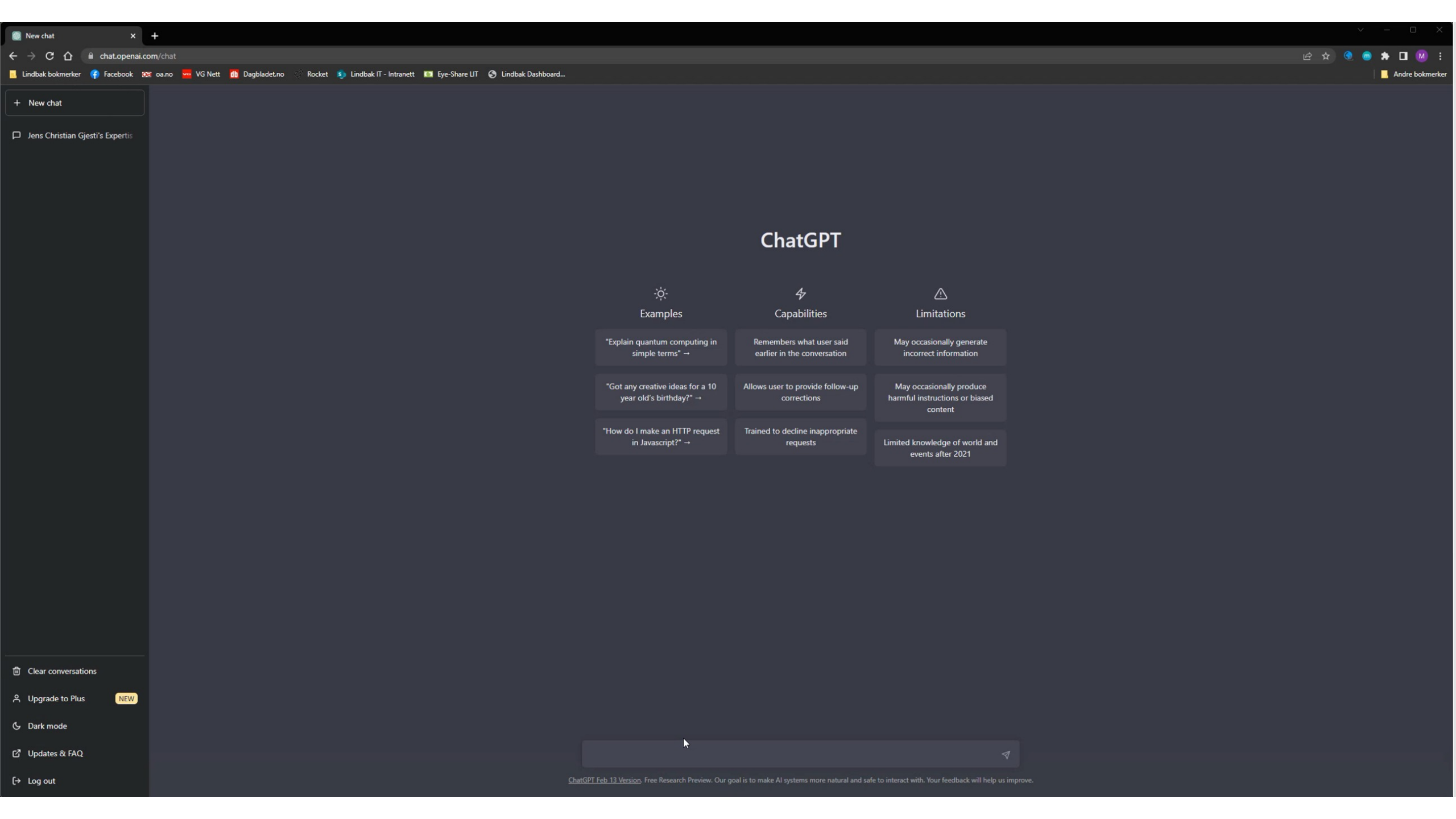
- Remembers what user said earlier in the conversation
- Allows user to provide follow-up corrections
- Trained to decline inappropriate requests



Limitations

- May occasionally generate incorrect information
- May occasionally produce harmful instructions or biased content
- Limited knowledge of world and events after 2021

Lag samboerkontrakt på 42 sekunder...



ChatGPT



Examples

- "Explain quantum computing in simple terms" →
- "Got any creative ideas for a 10 year old's birthday?" →
- "How do I make an HTTP request in Javascript?" →



Capabilities

- Remembers what user said earlier in the conversation
- Allows user to provide follow-up corrections
- Trained to decline inappropriate requests



Limitations

- May occasionally generate incorrect information
- May occasionally produce harmful instructions or biased content
- Limited knowledge of world and events after 2021



Can you generate some filler text for a historical novel? It would be a listicle from the French Third Republic outlining 12 regressive opinions held by a citizen



Sure, here are some examples of regressive opinions that a citizen of the French Third Republic might hold:

1. Women should not be allowed to vote or hold political office.
2. The French monarchy should be restored.
3. The Catholic Church should have more influence over the government.
4. The French Empire should be expanded through conquest and colonization.
5. The French language and culture should be promoted at the expense of other cultures and languages.
6. Slavery should be reinstated.
7. Jews and other minority groups should be denied equal rights.
8. Immigrants should be heavily restricted or banned altogether.
9. The government should censor the media and suppress dissenting views.
10. The French military should have more power and autonomy.
11. The economy should be based on protectionist trade policies and heavy industry.
12. Science and technology should be rejected in favor of traditional values and beliefs.



When did Queen Elizabeth II die?



Queen Elizabeth II is still alive. She is the longest-reigning current monarch in the world, having ascended to the throne in 1952. As of December 2021, she is 96 years old.



No, she died this year. I'm asking when that happened.



I apologize for misunderstanding your question. As I mentioned earlier, my training data only goes up until 2021, and I am not able to browse the web or access any information beyond that point. Therefore, I am unable to provide you with information about events that have occurred after 2021.



Chatgpt's bingo – 100% generert av

CHATGPT

url: <https://temp.no/bingo/>

B	I	N	G	O

Generate Number

javascript

Copy code

```
// Create an array to store the generated numbers
let numbers = [];

// Create a function to generate a random number between 1 and 75
function generateRandomNumber() {
    return Math.floor(Math.random() * 75) + 1;
}

// Create a function to generate a new number and add it to the numbers array
function generateNumbers() {
    // Generate a new number
    let number = generateRandomNumber();

    // Check if the number has already been generated
    if (numbers.includes(number)) {
        // If the number has already been generated, generate a new one
        generateNumbers();
    } else {
        // If the number hasn't been generated, add it to the numbers array
        numbers.push(number);
        // Update the HTML to display the new number
        document.getElementsByTagName("td")[number - 1].style.backgroundColor =
            "lightgreen";
        document.getElementsByTagName("td")[number - 1].innerHTML = number;
    }
}
```