



««NOE»» OM ««DATA»»

Arntzen de Besche
Advokat Tommy Dahlen
24.10.24

ARNTZEN

DE BESCHE

MARKEDSGRUPPE FOR TEKNOLOGI OG INNOVASJON



TOMMY



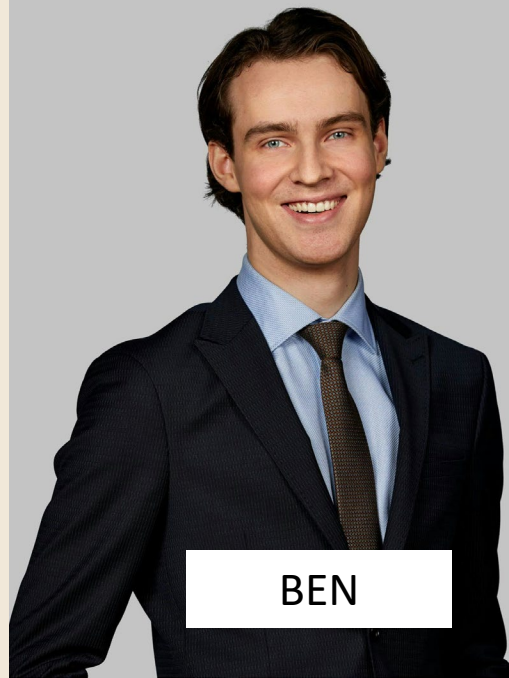
CHRIS



MARIANNE



VEIDE



BEN



RUNA

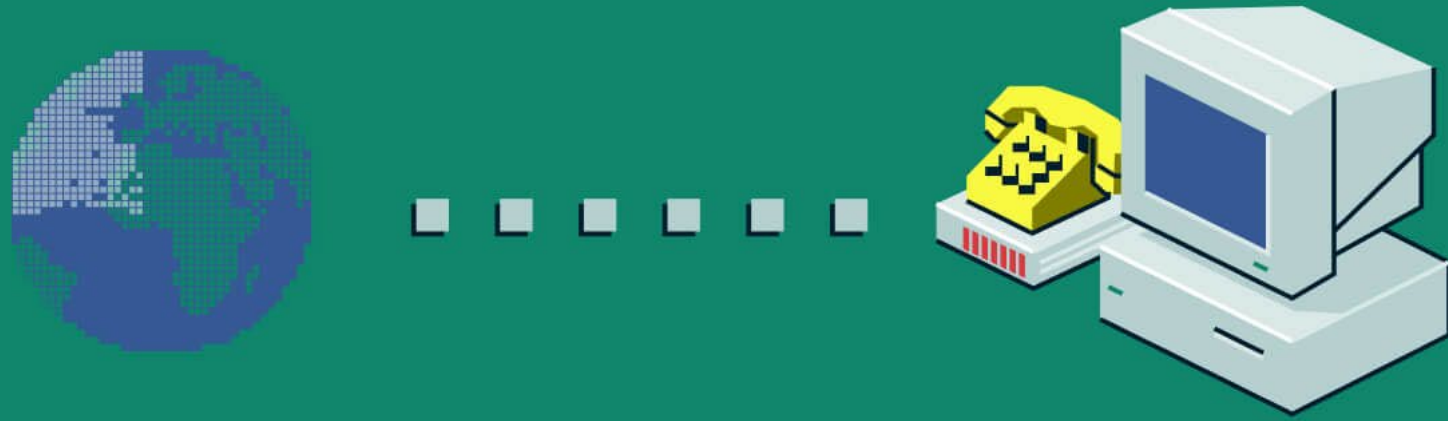


TRINE

DISCLAIMER

Gjennomgangen inneholder hallusinasjoner, avsporinger, anekdoter og misvisende/uriktige referanser for «dramatic effect»^{*}

^{*} it is not a bug, it is a feature



UNITED STATES DISTRICT COURT NORTHERN DISTRICT OF CALIFORNIA



WORLD HEAVYWEIGHT CHAMPIONSHIP ★ 15 ROUNDS



SHAWN FANNING ★ SEAN PARKER

METALLICA & DR. DRE

REIGNING RECORDING INDUSTRY CHAMPIONS

-VS-

NAPSTER

ONLINE DIGITAL FILE SHARING CHALLENGERS

ONLY ONE
WILL SURVIVE



LARS ULRICH ★ DR. DRE

★★★ \$94 MILLION AT STAKE ★ ASSETS WILL BE LIQUIDATED ★★★



vs

The New York Times

LITT OM BEGREPET «DATA»

ALL (FAKTISK) INFORMASJON
SOM KAN UTNYTTES I SEG
SELV ELLER I KOMBINASJON
SOM GRUNNLAG FOR
ANALYSE, DISKUSJON ELLER
BESLUTNINGER

EVOLUSJONSTRINN OG HISTORISKE FLASKEHALSER

S₁, E₁: FANGSTPROBLEM

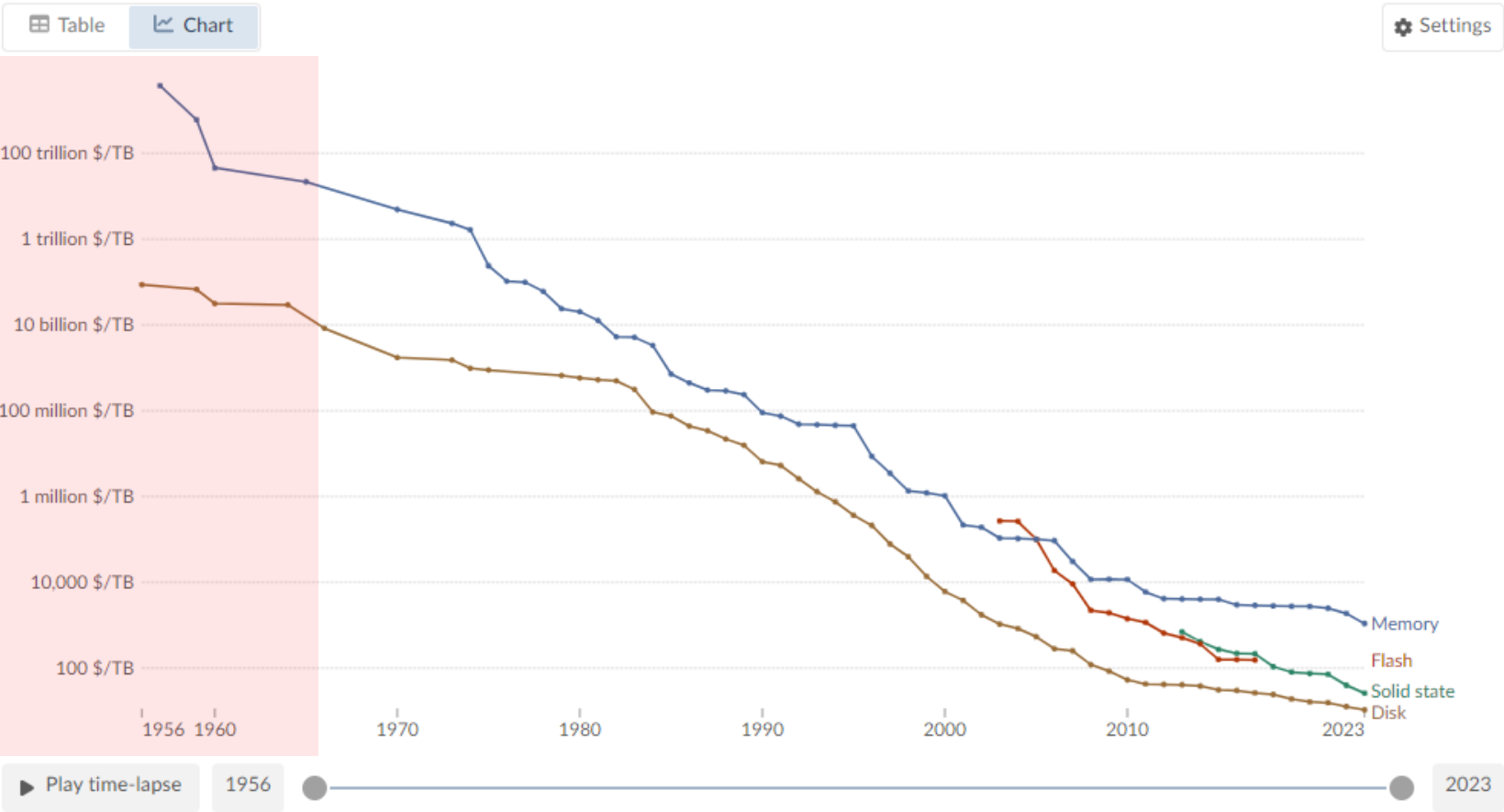
S₁, E₂: LAGRINGSPROBLEM

S₁, E₃: INNSIKTSPROBLEM

S₁, E₄: SIKKERHETSPROBLEM

Historical price of computer memory and storage

This data is expressed in US dollars per terabyte (TB), adjusted for inflation. "Memory" refers to random access memory (RAM), "disk" to magnetic storage, "flash" to special memory used for rapid data access and rewriting, and "solid state" to solid-state drives (SSDs).



Data source: John C. McCallum (2023); U.S. Bureau of Labor Statistics (2024) – [Learn more about this data](#)

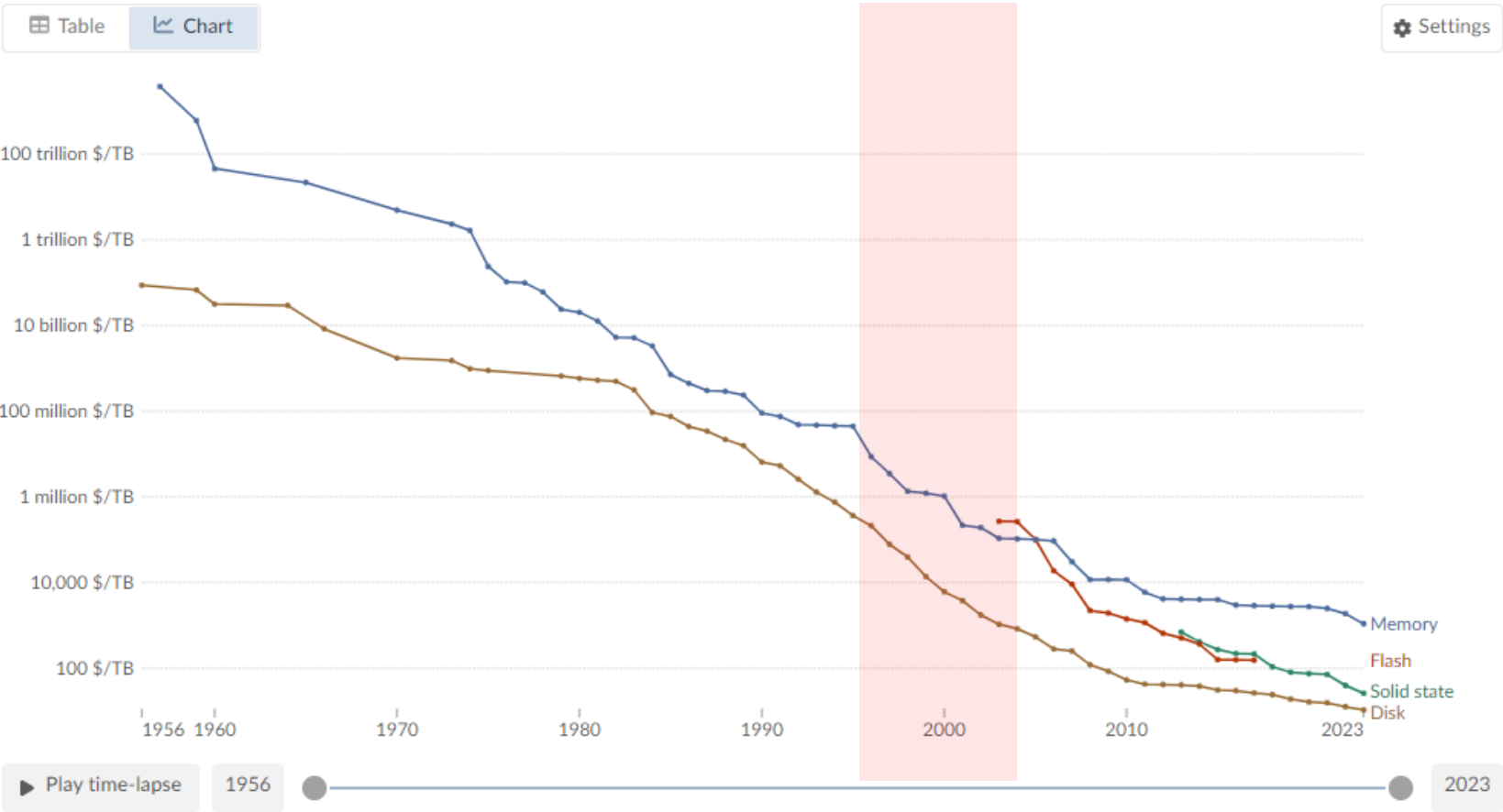
Note: For each year, the time series shows the cheapest historical price recorded until that year. This data is expressed in constant 2020 US\$.

OurWorldinData.org/technological-change | CC BY



Historical price of computer memory and storage

This data is expressed in US dollars per terabyte (TB), adjusted for inflation. "Memory" refers to random access memory (RAM), "disk" to magnetic storage, "flash" to special memory used for rapid data access and rewriting, and "solid state" to solid-state drives (SSDs).



Data source: John C. McCallum (2023); U.S. Bureau of Labor Statistics (2024) – [Learn more about this data](#)

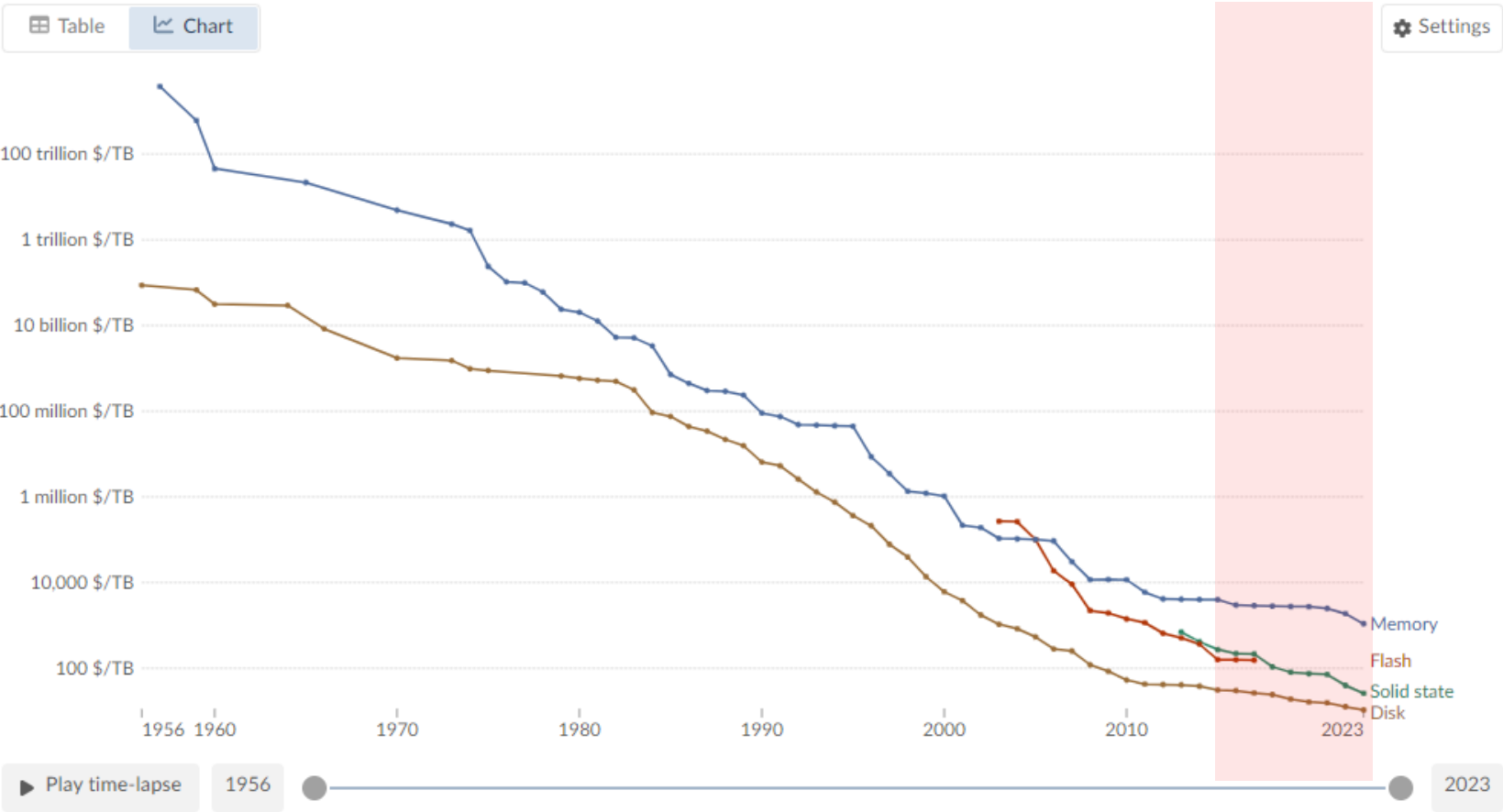
Note: For each year, the time series shows the cheapest historical price recorded until that year. This data is expressed in constant 2020 US\$.

OurWorldinData.org/technological-change | CC BY



Historical price of computer memory and storage

This data is expressed in US dollars per terabyte (TB), adjusted for inflation. "Memory" refers to random access memory (RAM), "disk" to magnetic storage, "flash" to special memory used for rapid data access and rewriting, and "solid state" to solid-state drives (SSDs).



Data source: John C. McCallum (2023); U.S. Bureau of Labor Statistics (2024) – [Learn more about this data](#)

Note: For each year, the time series shows the cheapest historical price recorded until that year. This data is expressed in constant 2020 US\$.

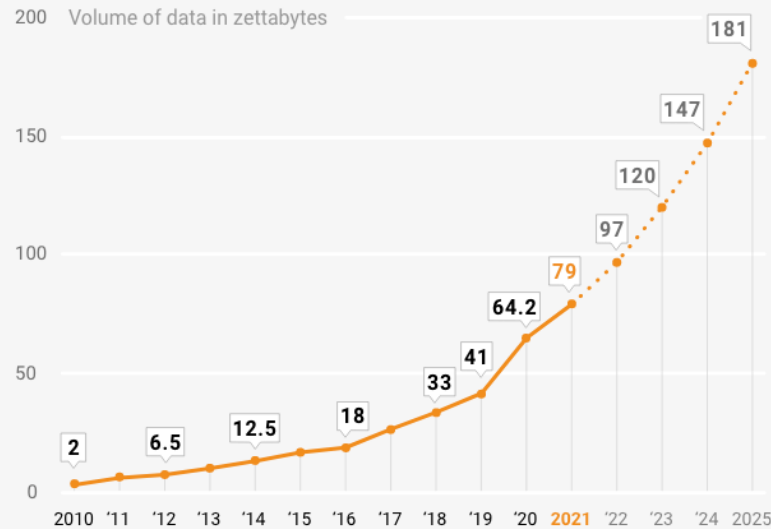
OurWorldinData.org/technological-change | CC BY



Volume of data created, captured, copied, and consumed worldwide



The volume of data generated, consumed, copied, and stored is projected to exceed 180 zettabytes by 2025

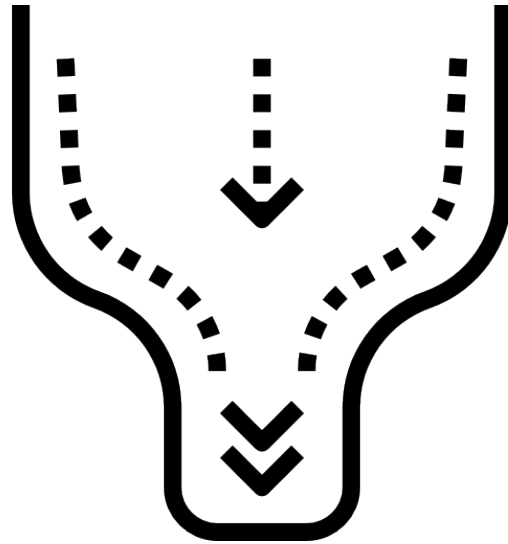


Source: statista.com



firstsiteguide.com

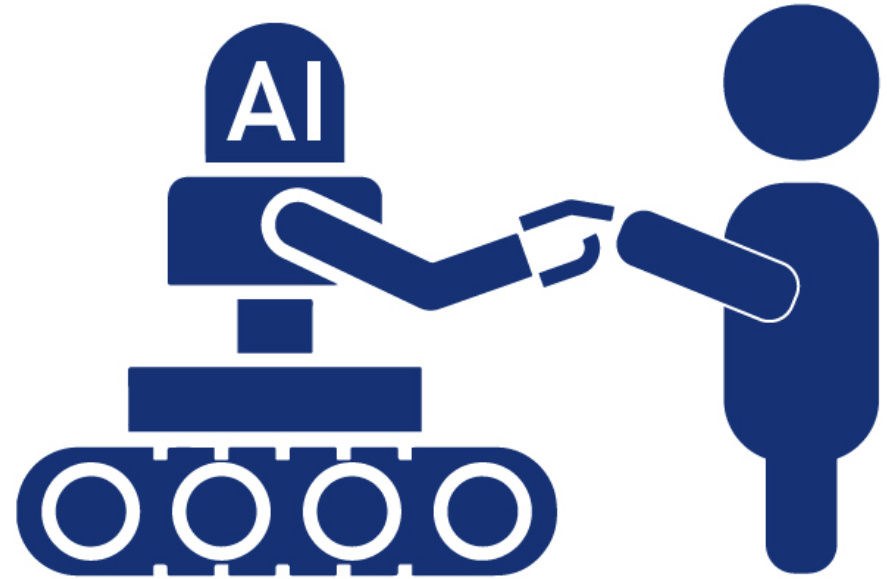
DATA



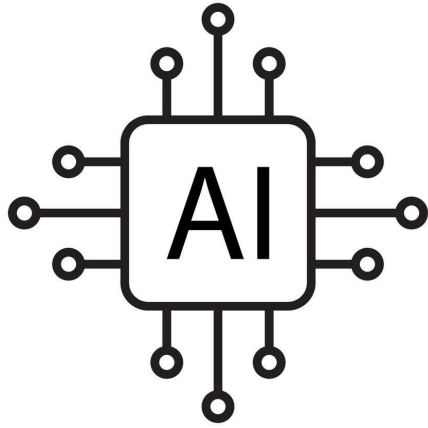
INNSIKT/HANDLING



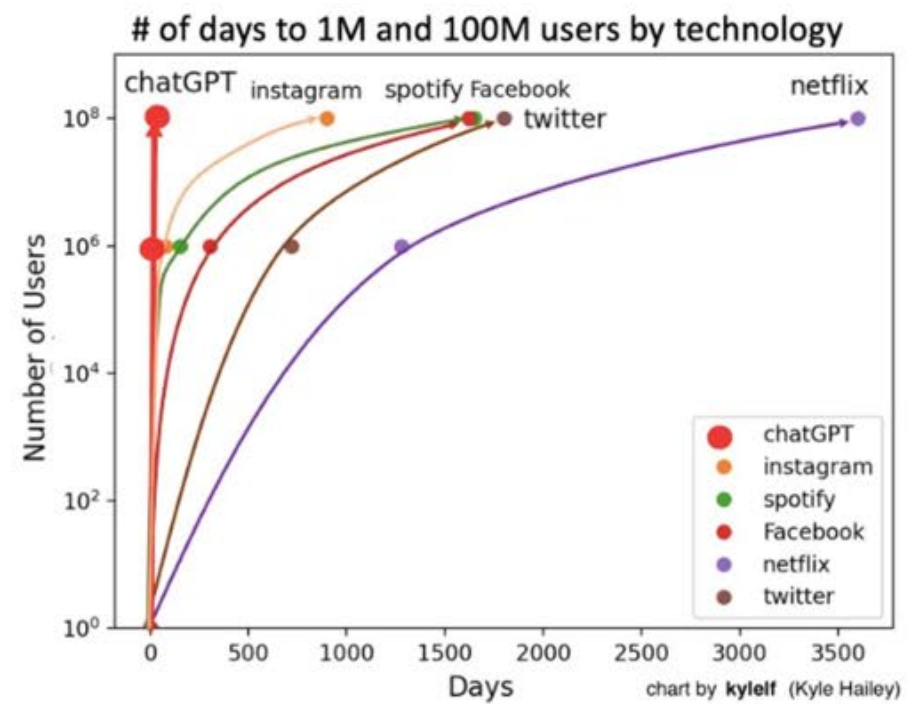
Industry 4.0
2010



Industry 5.0
2020



“a software-based system or machine that uses algorithms and data to simulate human intelligence processes. These systems can perform tasks **typically requiring human cognitive abilities....**”





FRA 100M TIL **200M** UKENTLIGE BRUKERE



1M VIRKSOMHETER – **80M** ENHETER



7 MRD USD I INVESTERINGER TIL UTVIKLING OG DRIFT FOR 2024



ESTIMERT **10MRD** USD FOR 2023/2024



Intelligent agents in AI will make your AI more useful

Today's AI models perform tasks such as generating text, but these are "prompted" — the AI isn't acting by itself. That is about to change with agentic AI, or AI with agency. By 2028, 33% of enterprise software applications will include agentic AI, up from less than 1% in 2024, enabling 15% of day-to-day work decisions to be made autonomously.

Intelligent agents in AI are goal-driven software entities that use AI techniques to complete tasks and achieve goals. They don't require explicit inputs and don't produce predetermined outputs. Instead, they can receive instructions, create a plan and use tooling to complete tasks, and produce dynamic outputs. Examples include AI agents, machine customers and multiagent systems.

PROBLEMSTILLINGER

HVORDAN SKAPE VERDI FRA
«**DATAFANGST**» VIA
«**INNSIKT**» TIL «**HANDLING**»

JURIDISKE
RAMMEBETINGELSER -
PERSONVERN,
SIKKERHETSKRAV M.V

**«THEY ARE EATING THE
CATS; THEY ARE EATING THE
DOGS; THEY ARE EATING THE
PETS OF THE PEOPLE THAT
LIVE THERE.....»»**



?



!

PROBLEMSTILLINGER

HVORDAN **SKAPE VERDI** FRA
«DATAFANGST» VIA
«INNSIKT» TIL «HANDLING»

JURIDISKE
RAMMEBETINGELSER -
PERSONVERN,
SIKKERHETSKRAV M.V

**VEIEN TIL VARIG
«VERDI»**

EIERSKAP TIL DATA

**KVALITET OG
ANVENDELIGHET** AV DATA

TILGANGSRETT TIL DATA

DISPOSISJONSRETT TIL DATA



VI ER **FOR MYE** FOKUSERT PÅ EIERSKAP OG
FOR LITE FOKUSERT PÅ TILGANG OG
DISPOSJONSRETT

**«EVERYTHING. EVERYWHERE. ALL AT
ONCE»**

EIERSKAP TIL DATA

PRIMÆRDATA/RÅDATA
(STAND-ALONE OG
DATABASER)

BEARBEIDEDE DATA
(«TRENINGSDATA»)

DERIVATER
(DATAMODELLER)

Men hva med **MODELLEN som sådan** – arkitektur og «veker»?

Vekt = matematisk uttrykk (numerisk) av relasjoner mellom «datapunkter»

Kan **MODELLEN** (summen av arkitektur og vekter) beskyttes?

Svaret må være «**JA**».

Modellen må kunne beskyttes som en database på samme måte som telefonkatalogen. Hvert datapunkt er ikke beskyttbar i seg – men sammenstilling av disse numeriske relasjonene derimot (og selv om modellene har lært relasjonene selv....).

TILGANGSRETT TIL DATA

EGEN TILGANG

ANDRES TILGANG

DISPOSISJONSRETT TIL DATA

**MULIGHETER OG
BEGRENSNINGER** MÅ
DEFINERES TYDELIGERE
- BÅDE FOR
EIER/BRUKER OG
TJENESTELEVERANDØR

“Upon full and final payment of all fees as outlined in this Agreement, all rights, title, and interest in and to any and all deliverables developed or created under the scope of this Agreement, including, but not limited to, **artificial intelligence (AI) models**, algorithms, software, data sets, and any related documentation or materials (collectively referred to as "Project Deliverables"), **shall vest exclusively in [Customer Name]**.”

Ownership of Customer Data

The **Customer shall retain exclusive ownership** and control over all data provided by the Customer or collected through the Customer's use of the Deliverables, consisting of **raw data, metadata, and training data** ("Customer Data"). The Supplier acknowledges that the Customer Data is confidential and proprietary to the Customer.

Supplier's Use of Derived Data

The Supplier shall have the right to freely use, modify, and develop insights, models, algorithms, and any other **derived data** from the Customer Data, including but not limited to **model weights and analytical insights**, provided that such derived data: a) **cannot** be reverse-engineered, deconstructed, or otherwise **used to identify, trace, or reasonably infer the identity of the Customer** or any individual associated with the Customer; b) **does not include any Customer Data in its raw form** or any data that would enable the identification of the Customer.

**BRUK AV «EGNE» DATA
SOM INPUT I
(GENERATIVE) AI-
VERKTØY**

MITT, DITT, VÅRT?

KONFIDENSIALITET

BREAKING

Samsung Bans ChatGPT Among Employees After Sensitive Code Leak

Siladitya Ray Forbes Staff

Covering breaking news and tech policy stories at Forbes.

Follow



May 2, 2023, 07:17am EDT

Updated May 2, 2023, 07:31am EDT

TOPLINE Samsung Electronics has banned the use of ChatGPT and other AI-powered chatbots by its employees, Bloomberg [reported](#), becoming the latest company to crack down on the workplace use of AI services amid concerns about sensitive internal information being leaked on such platforms.

KONFIDENSIALITET –tilsynelatende blitt et håndterlig problem («sandbox-modeller» i de fleste sentrale verktøy)

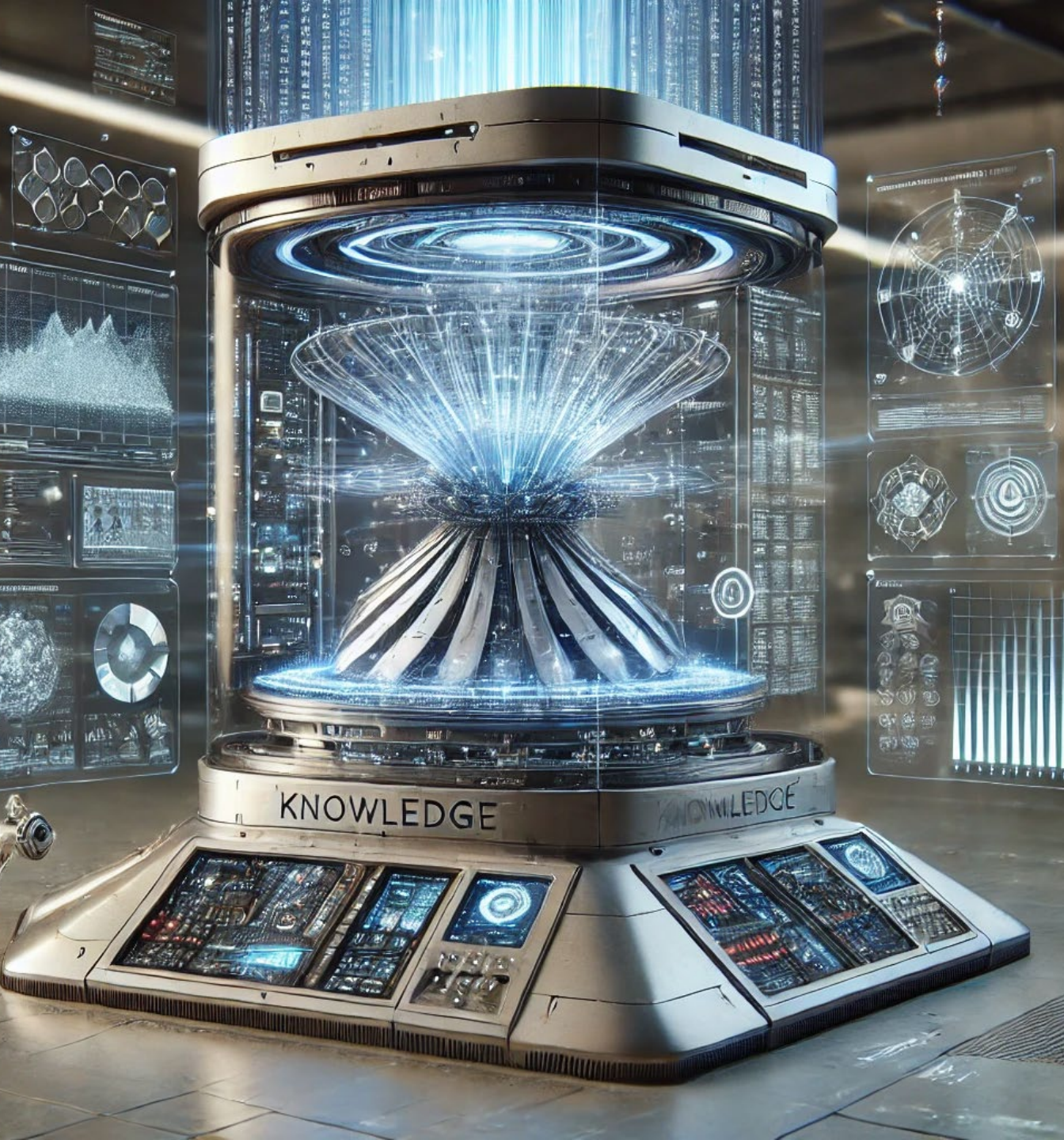
MEN...

No Use of Artificial Intelligence Systems

The Receiving Party agrees that **it shall not** use, process, input, upload, or otherwise expose any Confidential Information, in whole or in part, to any artificial intelligence **(AI) systems**, including but not limited to machine learning models, large language models (LLMs), data scraping tools, or any other automated systems or platforms that involve training, optimization, or inference from data, **whether or not the AI system retains, stores, or feeds back the input data into its models.**

Men hva med eierskap og kontroll med **OUTPUT** som vi gjør tilgjengelig for partnere, kunder og myndigheter?

Har en klar side mot <<**kontroll**>> som diskutert tidligere....



BUILDING THE KNOWLEDGE-MACHINE

Kloner av kollektiv
kunnskap og erfaring for å
bygge ny (derivert)
kunnskap

[nature](#) > [news](#) > article

NEWS | 30 August 2024

Researchers built an ‘AI Scientist’ — what can it do?

The large language model does everything from reading the literature to writing and reviewing its own papers, but it has limited applicability.

By [Davide Castelvecchi](#)



Fokus på vilkår som regulerer (overføring av) **eierskap til elementer** som inngår i AI-modeller, både algoritmer, verktøy, vekter og innhold/data.

Vurdere **restriksjoner** på kunder/partneres bruk av slike «deliverables» som treningsgrunnlag for store/små modeller



**BRUKEN AV «ANDRES
DATA»**

KONFIDENSIALITET

BRUDD PÅ **OPPHAVSRETT** –
UTØVELSE AV PATENTERT
TEKNOLOGI?

BRUDD PÅ
**LISENSFORPLIKTELSER OG
TJENESTEVILKÅR**

BETYDELIGE endringer på dette området siste 12 mnd.

EKSPLOSJON i eksplisitte restriksjoner på bruk av løsninger og innhold.

**OPPDATERTE TJENESTEVILKÅR – OPPDATERTE
TJENESTEVILKÅR – OPPDATERTE TJENESTEVILKÅR...**

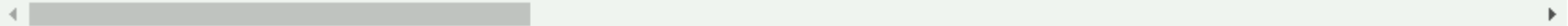
BY USING SCRIBD YOU AGREE NOT TO:

9.1 use Scribd for any purpose other than to receive original or appropriately licensed content, to add Descr

9.2 rent, lease, loan, sell, resell, sublicense, distribute, display or otherwise transfer the licenses grant

9.15 use any robot, spider, scraper, or other automated means to access Scribd, or copy, print, access, store

9.20 use any portion of the content on Scribd for the purposes of training a large language model.



**ETTERLEVELSESKONTROLL
OG KVALITET**

**TO BE OR NOT TO BE
(**RIGHT**): THAT IS THE
QUESTION**

MANGEL PÅ TRANSPARENS

Betydelige endringer i krav fra kunder og oppdragsgivere til **metoder, dokumentasjon og «etterprøvbarhet»** i resultater som er generert av/i AI-modeller

Results

Overall, 115 references were generated by ChatGPT, with a mean of 3.8 ± 1.1 per paper. Among these references, 47% were fabricated, 46% were authentic but inaccurate, and only 7% were authentic and accurate. The likelihood of fabricated references significantly differed based on prompt variations; yet the frequency of authentic and accurate references remained low in all cases. Among the seven components evaluated for each reference, an incorrect PMID number was most common, listed in 93% of papers. Incorrect volume (64%), page numbers (64%), and year of publication (60%) were the next most frequent errors. The mean number of inaccurate components was 4.3 ± 2.8 out of seven per reference.

Conclusions

The findings of this study emphasize the need for caution when seeking medical information on ChatGPT since most of the references provided were found to be fabricated or inaccurate. Individuals are advised to verify medical information from reliable sources and avoid relying solely on artificial intelligence-generated content.

How Is ChatGPT’s Behavior Changing over Time?

Lingjiao Chen[†], Matei Zaharia[‡], James Zou[†]

[†]Stanford University [‡]UC Berkeley

Abstract

GPT-3.5 and GPT-4 are the two most widely used large language model (LLM) services. However, when and how these models are updated over time is opaque. Here, we evaluate the March 2023 and June 2023 versions of GPT-3.5 and GPT-4 on several diverse tasks: 1) math problems, 2) sensitive/dangerous questions, 3) opinion surveys, 4) multi-hop knowledge-intensive questions, 5) generating code, 6) US Medical License tests, and 7) visual reasoning. We find that the performance and behavior of both GPT-3.5 and GPT-4 can vary greatly over time. For example, GPT-4 (March 2023) was reasonable at identifying prime vs. composite numbers (84% accuracy) but GPT-4 (June 2023) was poor on these same questions (51% accuracy). This is partly explained by a drop in GPT-4’s amenity to follow chain-of-thought prompting. Interestingly, GPT-3.5 was much better in June than in March in this task. GPT-4 became less willing to answer sensitive questions and opinion survey questions in June than in March. GPT-4 performed better at multi-hop questions in June than in March, while GPT-3.5’s performance dropped on this task. Both GPT-4 and GPT-3.5 had more formatting mistakes in code generation in June than in March. We provide evidence that GPT-4’s ability to follow user instructions has decreased over time, which is one common factor behind the many behavior drifts. Overall, our findings show that the behavior of the “same” LLM service can change substantially in a relatively short amount of time, highlighting the need for continuous monitoring of LLMs.

“However, the OpenAI o1 series of models is based on a new training and inference paradigm, which is specifically designed to address the types of problems present in the test sets used in the study.

In fact, by manually testing o1-preview with the example prompts described in the paper, a significant improvement is already observed on those problems where the study indicates that GPT4 fails.”





**“a greener, safer and
more understandable
AI for natural science
and technology”**

Professor emeritus Harald Martens (NTNU)



ARNTZEN
DE BESCHE

Tommy Dahlen

+47 994 57 460

tod@adeb.no