



SINTEF

KI og miljø – disse grepene bør du ta

Erik Johannes Husom

24.09.2026



Hvem er jeg?



Engineering Carbon Emission-aware Machine Learning Pipelines

Erik Johannes Husom
SINTEF
Oslo, Norway
erik.johannes.husom@sintef.no

Sagar Sen
SINTEF
Oslo, Norway
sagar.sen@sintef.no

Arda Goknil
SINTEF
Oslo, Norway
arda.goknil@sintef.no

ABSTRACT

The proliferation of machine learning (ML) has brought unprecedented advancements in technology, but it has also raised concerns about its environmental impact, particularly concerning carbon emissions. To address the imperative of environmentally responsible ML, we present in this paper a novel ML pipeline, named CEMAI, designed to monitor and analyze carbon emissions across the entire lifecycle of ML model development, from data preparation to training and deployment. Our endeavor involves an exhaustive evaluation process underpinned by three industrial case studies. These case studies are structured around the application of ML models to predict tool wear, estimate remaining useful lifetimes, and detect anomalies in the Industrial Internet of Things (IIoT). Leveraging sensor data originating from CNC machining and broaching operations, our research shows empirically the efficacy of carbon emissions as a dependable metric guiding the configuration of an ML development pipeline, enabling a balance between performance and environmental impact.

mitigate
tally res
in pers
perform
Our go
we engi
We st
with a l
environ
optimal
research
bon em
an open
based o
ated car
research
summit

The Price of Prompting: Profiling Energy Use in Large Language Models Inference

Erik Johannes Husom, Arda Goknil, Lwin Khin Shar, and Sagar Sen

1

Engineering Sustainable Agents: A Systematic Comparison of Agentic LLMs for Developer Workflows

MERVE ASTEKIN, SINTEF, Norway
YAN NAING TUN, Singapore Management University, Singapore
ARDA GOKNIL, SINTEF, Norway
ERIK JOHANNES HUSOM, SINTEF, Norway
LWIN KHIN SHAR, Singapore Management University, Singapore
HASAN SÖZER, Ozyegin University, Türkiye
RATNADIRA WIDYASARI, Singapore Management University, Singapore
HUI SONG, SINTEF, Norway

Large language models (LLMs) are increasingly used in software engineering, including agentic systems that coordinate multiple agents, but impose higher computational and environmental costs. In this paper, we present a comprehensive empirical study of agentic LLM systems across five software engineering tasks: code

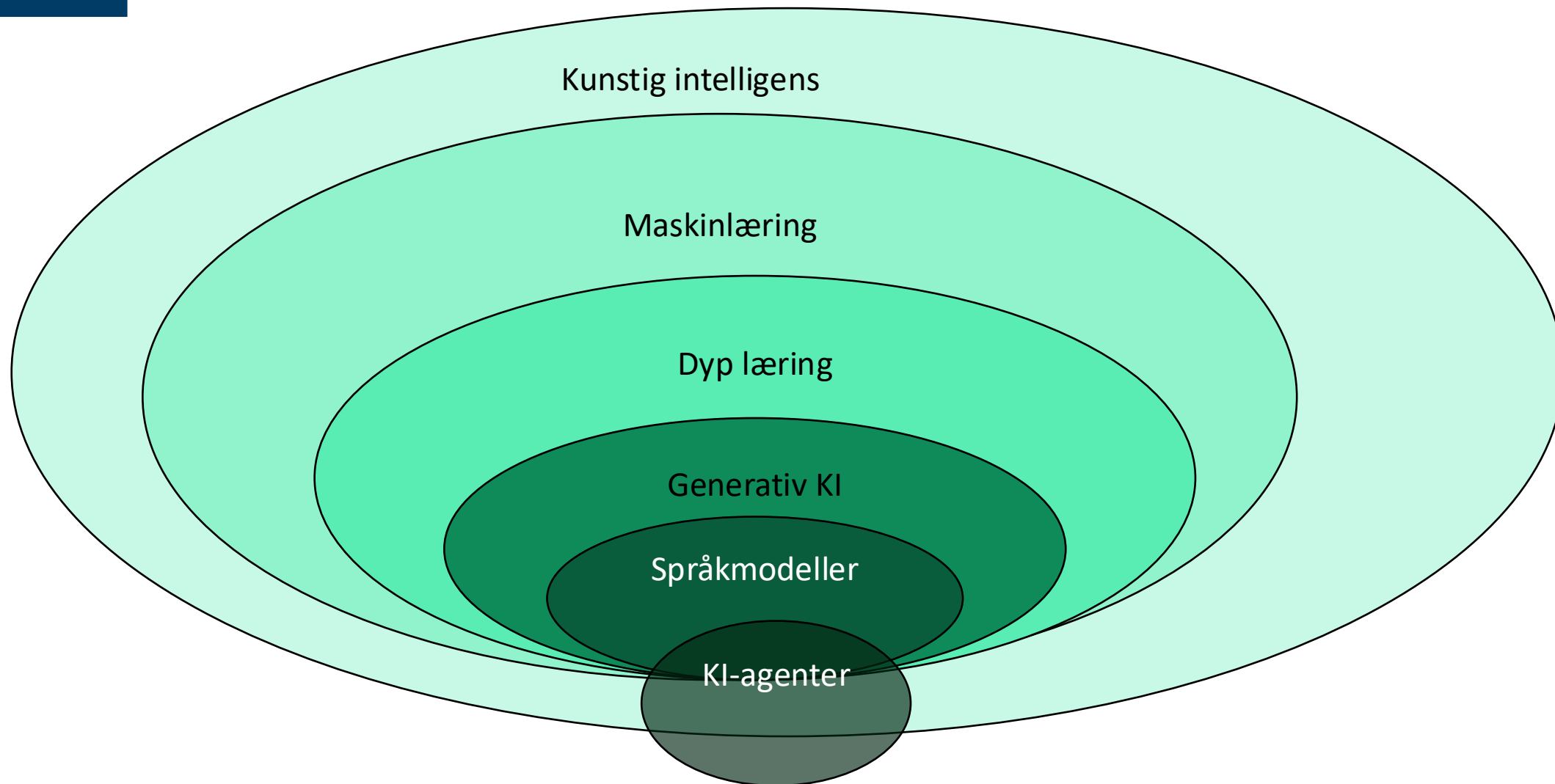
document review, thereby streamlining operations across various sectors. The ability of LLMs to generate human text with high accuracy and relevance has made them valuable tools for professionals, enabling them to focus on complex and creative aspects of their disciplines. **Context and Motivation.** As these models grow in operational complexity, their energy demands escalate accordingly [18]–[20]. The widespread deployment of LLMs across various applications, from code generation tools to iterative workflows to general-purpose chatbots used by hundreds of millions of users, makes individual usage a catalyst for massive resource consumption. Unlike training, which occurs once per model, inference is a continuous

Sustainable LLM Inference for Edge AI: Evaluating Quantized LLMs for Energy Efficiency, Output Accuracy, and Inference Latency

ERIK JOHANNES HUSOM, SINTEF Digital, Oslo, Norway
ARDA GOKNIL, SINTEF Digital, Oslo, Norway
MERVE ASTEKIN, SINTEF Digital, Oslo, Norway
LWIN KHIN SHAR, Singapore Management University, Singapore, Singapore
ANDRE KÅVSEN, Oslo Metropolitan University, Oslo, Norway
SAGAR SEN, SINTEF Digital, Oslo, Norway
BENEDIKT ANDREAS MITHASSEL, SINTEF Digital, Oslo, Norway
AHMET SOYLU, Kristiania University of Applied Sciences, Oslo, Norway and Seoul National University, Seoul, Republic of Korea

Deploying Large Language Models (LLMs) on edge devices presents significant challenges due to computational constraints, memory limitations, inference speed, and energy consumption. Model quantization has emerged as a key technique to enable efficient LLM inference by reducing model size and computational overhead. In this study, we conduct a comprehensive analysis of 28 quantized LLMs from the Ollama library, which applies by default Post-Training Quantization (PTQ) and weight-only quantization techniques, deployed on an edge device (Raspberry Pi 4 with 4 GB RAM). We evaluate energy efficiency, inference performance, and output accuracy across multiple quantization levels and task types. Models are benchmarked on five standardized datasets (CommonsenseQA, BIG-Bench Hard, TruthfulQA, GSM8K, and HumanEval), and we employ a high-resolution, hardware-based energy measurement tool to capture real-world power consumption. Our findings reveal the trade-offs between energy efficiency, inference speed, and accuracy in different quantization settings, highlighting configurations that optimize LLM deployment for resource-constrained environments. By

Ulike former for KI





SINTEF

Mål for virksomheter

1. Øke andel KI-prosjekter som måler ressursbruk i utvikling eller drift
2. Øke andel innkjøpte KI-tjenester der leverandøren dokumenterer miljørelevant informasjon
3. Øke andel innkjøpte KI-tjenester der modell (eller region) er valgt ut fra behov og miljøhensyn
4. Øke andel ansatte som har fått opplæring i ressurseffektiv bruk av KI



SINTEF

Eksempler på tiltak for virksomheter

- Velg mindre modeller eller lavere tjenestenivå når det dekker behovet
- Vurder om KI er nødvendig før dere bygger eller kjøper en løsning



SINTEF

Kostnad vs nytte

- Alt har en kostnad – når er nytten verdt kostnaden?
- Hvordan måler vi nytteverdi?
- Hvilke KPI-er (Key Performance Indicators) står i fare for å kortsluttes?



SINTEF

Blindvei: Mennesker må kuttes ut

scientific reports

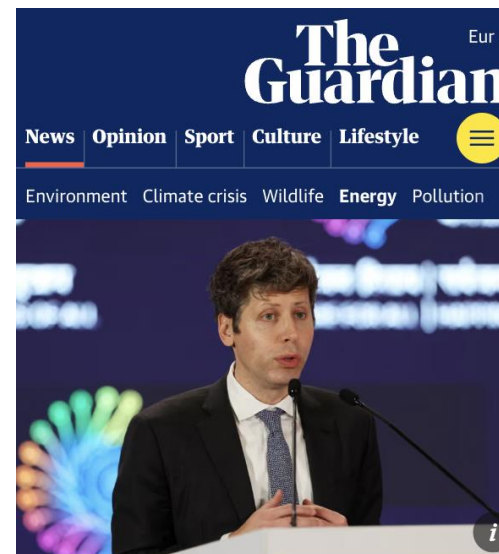
OPEN

The carbon emissions of writing and illustrating are lower for AI than for humans

Bill Tomlinson^{1,2✉}, Rebecca W. Black¹, Donald J. Patterson^{1,3} & Andrew W. Torrance^{4,5}

As AI systems proliferate, their greenhouse gas emissions are an increasingly important concern for human societies. In this article, we present a comparative analysis of the carbon emissions associated with AI systems (ChatGPT, BLOOM, DALL-E2, Midjourney) and human individuals performing equivalent writing and illustrating tasks. Our findings reveal that AI systems emit between 130 and 1500 times less CO₂e per page of text generated compared to human writers, while AI illustration systems emit between 310 and 2900 times less CO₂e per image than their human counterparts. Emissions analyses do not account for social impacts such as professional displacement, legality, and rebound effects. In addition, AI is not a substitute for all human tasks. Nevertheless, at present, the use of AI holds the potential to carry out several major activities at much lower emission levels than can humans.

Check for updates



Sam Altman

This article is more than 6 months old

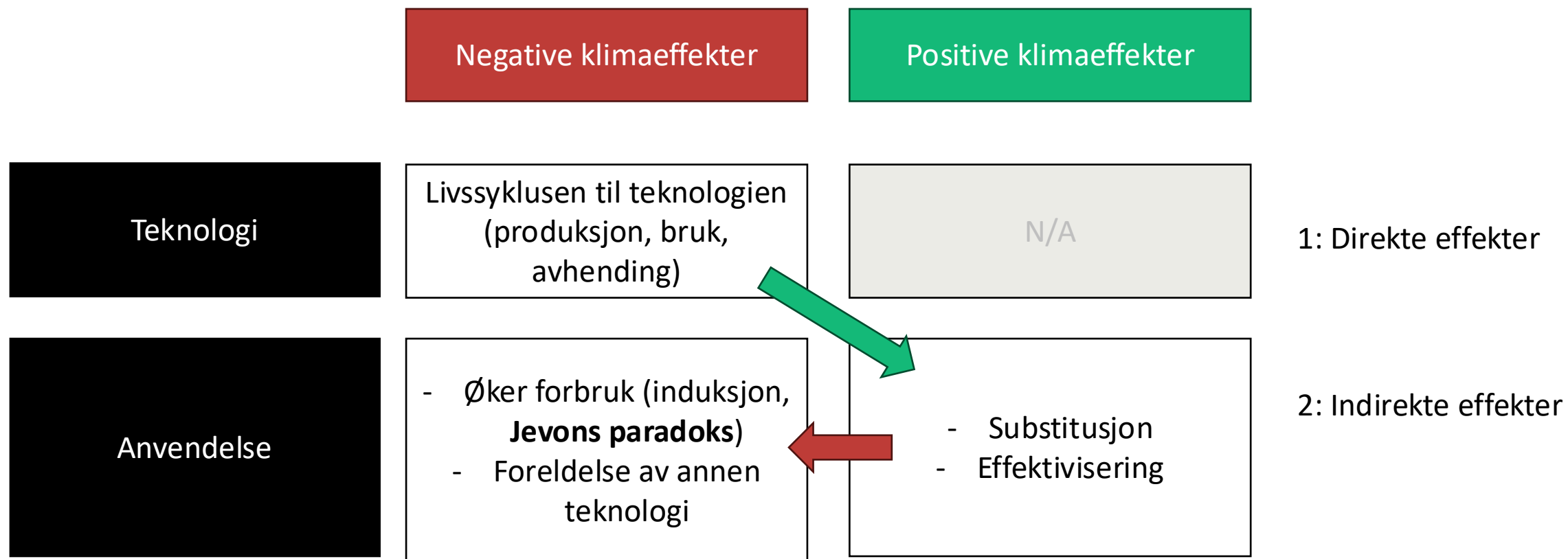
Sam Altman defends AI's energy toll by saying it also takes a lot to 'train a human'

OpenAI CEO also downplayed concerns about how much water datacenters require at AI summit in India

Eric Berger

Mon 23 Feb 2026 18:40 CET

Tilbakekoblingseffekter ved bruk av KI



Tilpasset fra: L. Hilty & B. Aebischer (ed.), « ICT for Sustainability: An Emerging Research Field » (Chap. 1), in ICT Innovations for Sustainability, Springer, 2015]



Klimaregnskap for KI

Tiltak for individer/virksomheter	Vanskelighetsgrad	Kommentar

Klimaregnskap for KI

Tiltak for individer/virksomheter	Vanskelighetsgrad	Kommentar
Redusere det direkte klimaavtrykket til KI-bruk for virksomheter	Enkelt (f. eks., token-begrensning)	Vanskelig å begrense bruk i en fase hvor utforskning er nødvendig.
Redusere totalt klimaavtrykk som følge av KI-bruk for virksomheter	Vanskelig (Jevons paradoks, endring i atferd, etc)	Behøver analyse av bruksområder og indirekte effekter.
Redusere totalt klimaavtrykk som følge av KI-bruk globalt	Ekstremt vanskelig	Investeringene i teknologien er ekstremt høye, med få/ingen bevis på at ressursbruken fører til lavere klimautslipp.



SINTEF

Tredemøllekappløpet

“Now, *here*, you see, it takes all the running you can do, to keep in the same place”

— Lewis Carroll, *Through the Looking-Glass*



SINTEF

Miljøfokus har andre positive effekter

“Open source/open-weights”-modeller gir:

- Bedre oversikt over kostnader
- Mer forutsigbarhet og stabilitet
- Mindre bruk-og-kast av modeller
- Digital suverenitet / “resilience”



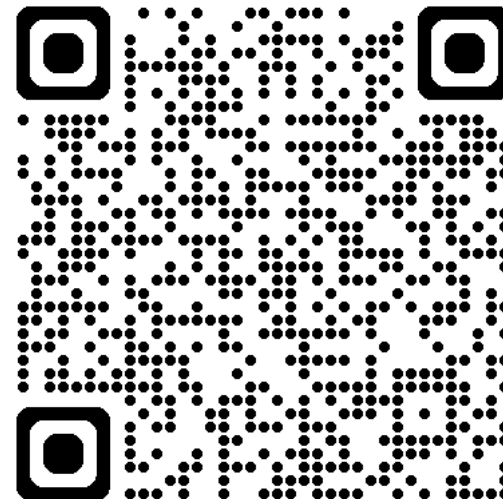
SINTEF

KI og miljø – hvilke grep bør du ta?

- Skaff overblikk over ressurskostnader og miljøavtrykk
- Definer konkret hva KI-verktøyene skal brukes til
- Vurder kostnaden opp imot nytten
- Tenk over indirekte effekter av KI-bruk

Technology for a better society

erik.johannes.husom@sintef.no



QR-kode til presentasjon



SINTEF

Power Hungry Processing: ⚡ Watts ⚡ Driving the Cost of AI Deployment?

ALEXANDRA SASHA LUCCIONI and YACINE JERNITE, Hugging Face, Canada/USA

EMMA STRUBELL, Carnegie Mellon University, Allen Institute for AI, USA

