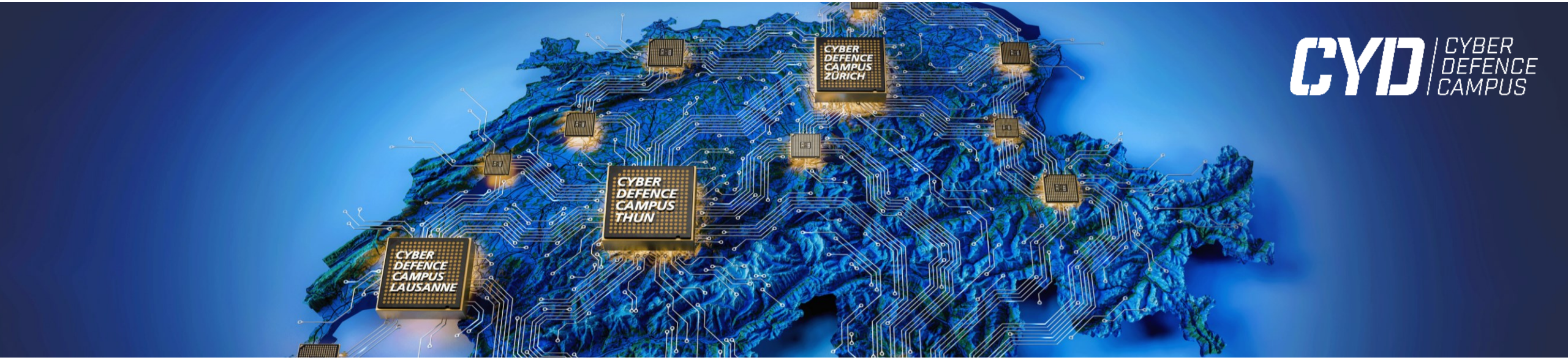
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra



Superintelligence

An Initial Assessment and Trend Analysis

Monday, June 16, 2025

Alexandre Vallotton & Valentin Mulder, Cyber-Defence Campus

From Today's AI to Superintelligence

Feature	Artificial Narrow Intelligence (ANI)	Artificial General Intelligence (AGI)	Artificial Super-intelligence (ASI)
Scope	Single Task	General tasks (human level)	All tasks (beyond human)
Learning adaptability	Minimal	High	Extrem
Consciousness	None	Possibly	Possibly
Current Existence	Already here	Not yet (R&D)	Theoretical
Control Challenge	Low	Medium	High

"**Superintelligence** represents a **theoretical** form of AI that surpasses human cognitive abilities **across all domains**. Unlike today's Narrow AI systems that excel at specific tasks, **Superintelligence would demonstrate extreme learning adaptability and potentially pose significant control challenges**"

Cole Stryker Tim Mucci. IBM. What is artificial superintelligence? (2023)
IBM. What is strong ai? (2021)

Key Technologies and Challenges

ANI (Achieved)

Key Technologies:

- Deep Learning
- Reinforcement Learning
- Large Language Models (LLMs)

1

2

AGI (R&D Phase)

Key Technologies:

- Neurosymbolic AI Transfer & Meta-Learning
- Cognitive Architectures Autonomous
- Self-Improving Systems

Key Challenges:

- Scalability of architectures
- Casual reasoning
- Robust long-term memory

3

ASI (Theoretical)

Key Technologies:

- Recursive Self-Improvement
- Neural Architecture Search
- Quantum AI
- Brain-Computer Integration

Key Challenges:

- Control & Alignment
- Ethical & Existential Risks
- Energy & Compute Constraints

Cole Stryker Tim Mucci. IBM. What is artificial superintelligence? 2023.
IBM. What is strong ai? (2021)

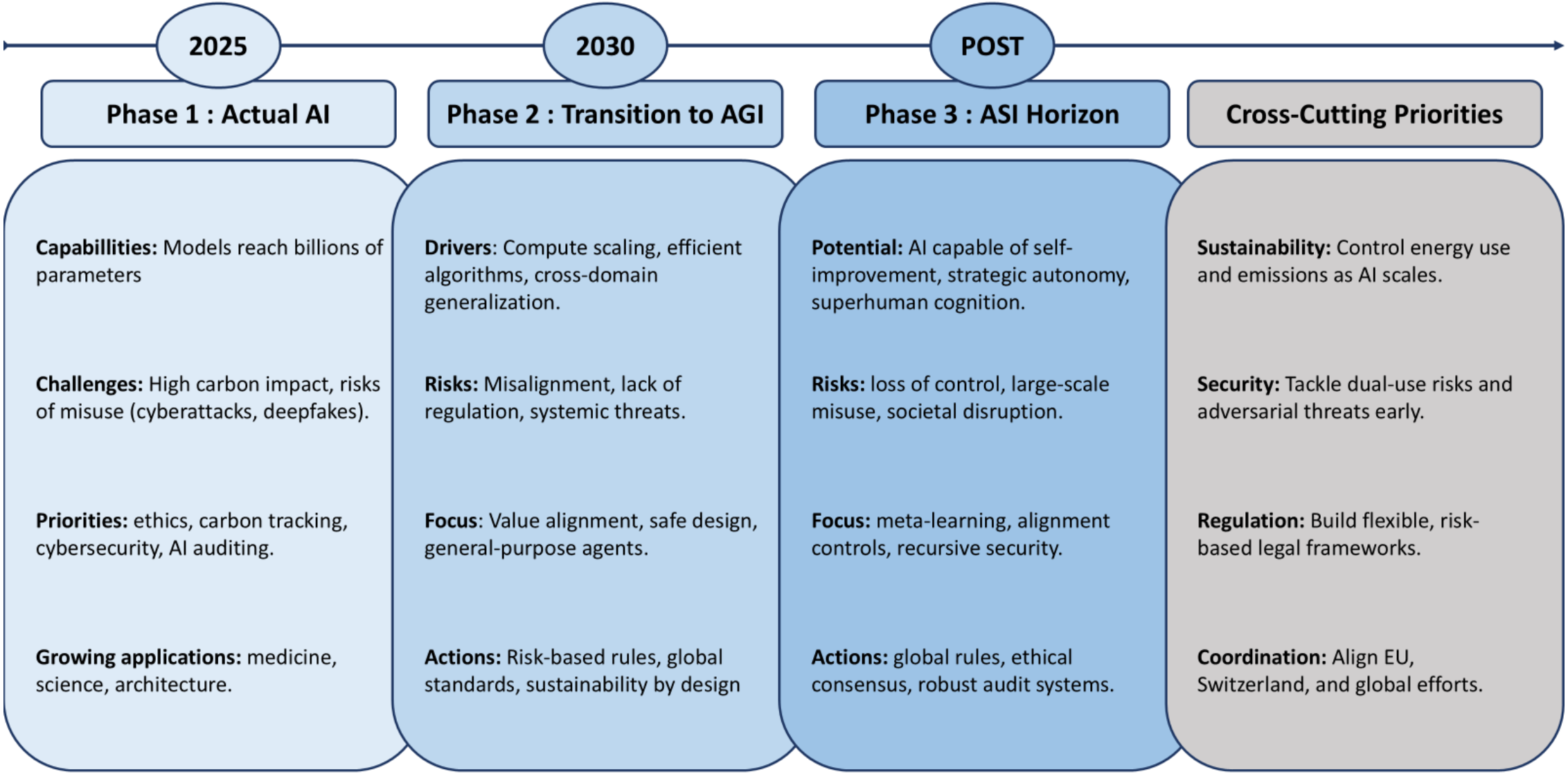
Comparative Analysis of Leading LLMs

Technology	Grok-3	Gemini 2.0	ChatGPT (GPT-4-turbo)	DeepSeek-R1	Claude 3
1. Core Architecture					
Transformer Architecture	X	X	X	X	X
(Sparse) Mixture-of-Experts ((S)MoE)	X (suspected)	X	X (suspected)	X	X
Multi-Head Latent Attention (MLA)	X		X (likely)	X	
Double-pass Architecture	X (likely)				
2. Optimization Techniques					
Multi-token Optimization	X	X	X (likely)	X	
Multi-round Optimization		X		X (suspected)	
KV Cache Compression			X	X	X
Hybrid RL-SFT	X (hybrid)		X (hybrid)		X
RL Without SFT				X	
Distillation to Smaller Models		X	X	X (Qwen-32B, Llama-3.1-8B)	
3. Learning & Training Paradigms					
RL	X	X	X	X	X
RLHF	X	X	X	X	X
Self-Correcting RL (SCoRe)		X (suspected)			
Self-Supervised Learning (SSL)	X	X	X	X (suspected)	X
4. Knowledge Integration & Reasoning					
Retrieval-Augmented Generation (RAG)	X (suspected)	X	X (suspected)	X (likely)	X (suspected)
Modular Model Approach	X	X			X (Haiku, Sonnet, Opus)
Built-in Error Detection and Correction	X (suspected)		X (likely)		
Metacognitive Abilities					X
5. Safety & Alignment					
Constitutional AI Framework					X
Censorship				X (Political censorship)	
RLHF	X	X	X	X	X
SCoRe		X			
6. Capabilities & Interfaces					
Multimodal Capabilities	X (suspected)	X (Text, Image, Audio, Video)	X (Text, Image, Voice)	X (suspected)	X (Text, Image)
Extended Context Window		X (>30K tokens)	X (128K tokens)		X (200K tokens)
Scalability Beyond One Million Tokens					X (Potential)

Ege Erdil. How has deepseek improved the transformer architecture? (2025)

"Joel Wembo". "deepseek vs. openai vs. grok 3 — a tech saga technical comparison". "Medium", (2025)

Roadmap to ASI - Trends & Forecasts



Nestor Maslej, Loredana Fattorini, and al. “The AI Index 2025 Annual Report,” AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2025.

Potential Threats, Risks and Limitations of Advanced AI systems

Global prohibited practices:

- **Cognitive manipulation** to manipulate human vulnerabilities
- **Mass biometric surveillance** and **Social scoring**
- **Autonomous decision-making** in justice, law enforcement, sensitive contexts

Other threats to consider:

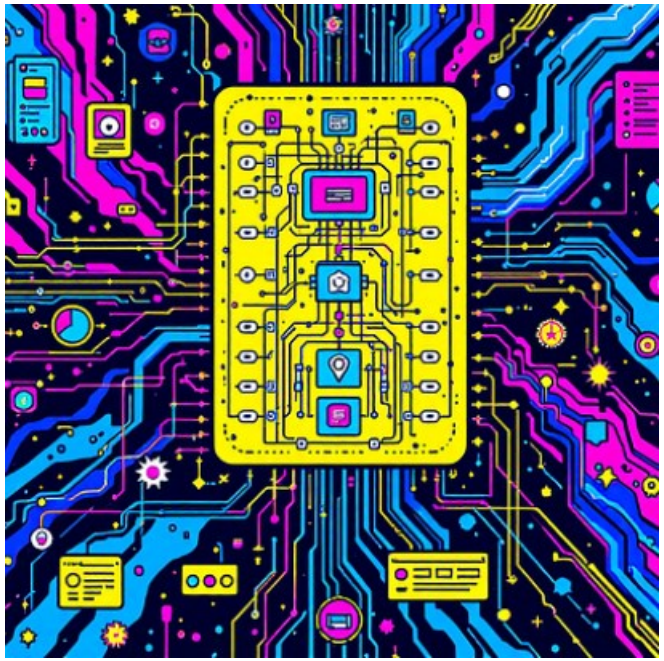
- **Concentration of power**
- **Strategic misuse**
- Economic disruption

European Parliament and Council of the European Union. Regulation (EU) 2024/1689: The Artificial Intelligence Act. 2024. Official Journal of the European Union
Allan Dafoe. Ai governance: a research agenda. Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK, 1442, 1443., 2018.

Highlights of Major Global Players

Strategy	Private sector innovation & investment	Rapid industrialization Integration of AI	Collaborative approach Academic and Industrial
Number of AI models (2023)	**>60 major** Machine Learning Models	**15 major** Machine Learning Models	EU will **overtake China** in the number of AI models
Fundings (2023)	**>$67 billion** investments	**~$7.8 billion** investments	**>$8 billion** investments
Other informations	**Leading** the World in AI research	**50%** of global **industrial robot** installations	Increase in public investment by **67 times**

Nestor Maslej, Loredana Fattorini, and al. "The AI Index 2024 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2024.

Conclusion

- 10 printed copies of the full report are available today.
- The online version will be released soon (currently undergoing internal approval).
- The **Technology Monitoring team** will **continue with ongoing monitoring** efforts.