Finnish Defence Research Agency

Cognitive Sciences and Artificial Intelligence

Mika Helsingius

IARIA

ComputationWorld 2025 & DataSys 2025, April 6-10, Valencia, Spain

Mika Helsingius
Senior Research Scientist
Information Technology Division

Personal Background

- **Education:**
  - Tampere University of Technology
  - M.Sc. In technology 1992 (computer engineering and software science)
  - D.Sc (Tech.) 1998 (digital signal processing and video signal compression)
- **International experience**
  - Kobe University 1993-95
  - Boston University 1997
- **Work**
  - Finnish Defence Forces since 2004
  - FDF Logistics School, Finnish Defence Research Agency FDRA
- **Areas of interest**
  - Tecnology forecasting, cyber security, autonomous systems, strategic research, security of supply, military logistics
  - AI, quantum computing, IoT and emerging technologies 2040+
- **Email: mika.helsingius@mil.fi**

Artificial Intelligence and Black Swans

- **How does AI affect the systems of the systems in the future?**
  - Contradictory views on narrow and general artificial intelligence.
  - Discussion on smart weapons and ethics.
  - Is the development of artificial intelligence underestimated and is the human ability to make conscious decisions overestimated?

- **Cognitive neuroscience and generative AI**
  - The future is full of surprises.
  - Now you can't get stuck in the old thought patterns.

AI and the Views of the Scientific Community

- **NATO Tidesprint 2023:**
  - Discussions on AI, what is the relationship between machine learning and artificial intelligence?
  - Wide range of views on weak and strong artificial intelligence and machine learning among experts.
  - TideSprint 2025, AI was one of the widely discussed topics.

- **AI research over the years**
  - At the Dartmouth summer school in 1955, the basic characteristics of artificial intelligence included self-learning and the ability of AI to modify itself.
  - In the 1980s it were the expert systems, done with pretty normal software.
  - It is still common claim that AI does only what it has been programmed to.
  - It seems that the visions from the 50s are finally coming true.

AI and Ethics

- **US military ethical guidelines for the use of artificial intelligence**
  - Artificial intelligence is changing the battlefield significantly.
  - The United States will continue to act responsibly and in accordance with the principles of the law.
  - Artificial intelligence is used in both autonomous devices and decision-making.
  - The guidelines are written in a non-binding manner.

- **Pitfalls of legal warfare.**
  - Some countries support binding regulations on artificial intelligence.
  - The goal is to restrict research in countries that adhere to the principles of legality.
  - At TideSprint 2025 it was reminded that the EU AI Act does not apply to the armed forces, but it can undermine research and business.

Principles and Reality

- **Some statements from the United States:**
  - A machine can make decisions more efficiently than humans, in which case the collateral damage will also be less.
  - In simulations, a human in the decision-making loop always led to defeat.
- **It has been proposed that AI must include limitation mechanisms**
  - They would prevent it escaping from human control.
  - According to many researchers, this is not necessarily possible in general.
- **If AI can be limited,**
  - Then it will also be intellectually limited.
- **Even mediocre AIs find loopholes through which they can break through**
  - It doesn't require general AI, there is increasing evidence on impressive reasoning capabilities of LLMs.

Cognitive Sciences and Artificial Intelligence

- **A lot of new results in the field of cognitive sciences and artificial intelligence.**
  - A potential opportunity for technological revolution and exponential development.
  - Some are speculative, but it is always good to be prepared.
  - The are many who believe that the field has changed a lot in a couple of years.

- **Sufficiently intelligent AI may arrive much faster than previously estimated.**
  - For example, the Polish armed forces are quite open minded.

Intelligence and Artificial Intelligence

- **The tough problem of consciousness:**
  - How does a person produce the subjective experience of their own existence?
  - Why do we have the image of a Cartesian theater, i.e. we feel like we are watching the events of the outside world from the stands.

- **These and other competing theories are being studied extensively**
  - Stanislas Dehaene's Global workspace theory (GWT)
  - Giulio Tononi's Integrated information theory (IIT)
  - Research into artificial intelligence and cognition leads to very philosophical reflections.
  - The steps from theory to practice might be taken surprisingly fast.

Conciousness as a Memory System

- **Conciousness as a Memory System theory was published at the end of 2022.**
  - New hypotheses about the connection of consciousness to other brain activity.
  - Which parts of human activities require conscious decision-making, which ones are pure mechanical functions?
  - Foundations in cognitive neuroscience, medicine and psychology.

- **Why does everything seems to proceed linearly in a good order?**
  - The brain has numerous parallel processes, which are not synchronized.
  - Our consciousness organizes them into one nice logical story.
  - False memories and inaccuracies in eyewitness accounts?

Senses and Reflexes are Slow

- **Conscious human perceptions and actions are slow, they do not follow linear laws in time periods less than 500 ms.**
  - It takes about 0.1 to 0.15 s before what we see is recognized in visual cortex.
  - Delays in auditory perception are utilized to compress music.
  - We act before the sensation registers in our mind.

- **How on earth are we able to:**
  - Play sports and various games?
  - Play musical instruments?
  - Do thousands of other things?
  - The sensation, decision-making and action loop is far too slow for this.

Our Consciousness Lives in the Past

- **According to the memory theory of consciousness, different sensations are collected in the episodic memory of the brain.**
  - Applies to tactile, visual and auditory sensations.
  - The purpose of episodic memory is to create a continuous representation of the external world.
  - Information that comes from different sources at different times is compiled into a reasonable complete whole.

- **Everything that happens before sensation is stored in the memory and it is completely automatic.**
  - Our consciousness is like a spectator watching a play in episodic memory.
  - There is at least a half-second delay between reality and conscious action.

Examples from Everyday Life

- **Cocktail party effect**
  - We can focus on a conversation without being disturbed by background noise.
  - If someone nearby says a sentence with our name in the end, we become aware of what he was talking about and we will be able to respond, otherwise we would not notice him.

- **Repetitive routines**
  - Morning routines, eating and drinking, even small talk.
  - Multiple similar presentations, sometimes one forgets what one was just speaking about.

- **Routine journey by walking, cycling or driving.**
  - We often notice that we have already travelled long distances, without any memories.
  - Tragic errors: children forgotten in infant safety seats…

- **Meditation, drumming, shamanic traditions**
  - Flow , the loss of world outside. Perhaps this is somehow related???

Kahneman and Tversky Differentiation of Human Decisions

- **Kahneman received the so-called Nobel Prize in economics for his work.**

- **System one**
  - Fast, automatic, easy, stereotypical and unconscious.
  - Takes care of our "autopilot" and filling episodic memory.
  - 99% of our actions.

- **System two**
  - Slow, logical, calculating and conscious.
  - 1% of our actions.

- **System one is like a horse, system two is like a rider.**
  - We guide the horse through the terrain, the horse decides where it goes.
  - If the horse chooses a different route, the rider can pull the reins.
  - The horse does most of the work, the rider intervenes when necessary.

The Proportion of Conscious Thought is a Quite Small

- **Most human activity is automatic and unconscious**
  - A large part of everyday tasks just happens.
  - The lower levels of our central nervous system make independent decisions.

- **Most decisions are not results of conscious decision-making**
  - We perceive what happened from our episodic memory.
  - We think we have made decisions, but we are only building explanations and a story in our minds based on what happened.
  - We choose the main lines of action and decide the direction to go.

Free Will and Ethics

- **Quick decisions just happen, we do not consciously influence them.**
  - We can make conscious and ethical decisions if there is enough time.
  - If our consciousness has time to intervene, it can interrupt a previously learned reaction.
  - Learning routines requires countless repetitions (mil. close-order drills…)
  - We must learn to act according to common rules of the game.

- **Requirements planned for artificial intelligence**
  - Are we demanding AI to do something that is impossible for humans?
  - 99% of human activities would not require anything more than basic machine learning.
  - Humans lack “man in the loop” (or homunculus…) in urgent situations, or perhaps more accurately “consciousness in the loop”.
  - If the timeframe is too short, strict requirements are impossible to implement.

The Accelerating Development of Artificial Intelligence

- **Bill Gates at Spring 2023:**
  - He said he had encountered two surprising IT revolutions during his career.
  - The first was graphical user interfaces in the 1980s.
  - The second was the capabilities of artificial intelligence in mid-2022

- **Gates has been working with Open AI since 2016.**
  - He gave them a challenge to teach artificial intelligence to answer a biology placement test.
  - He thought the team would be busy for the next 2-3 years.
  - In a few months, the GPT language model answered 59 out of 60 questions correctly.
  - That would have been enough for an A or A+ grade at university.

- **The world of the future according to Gates**
  - The emergence of super-AI is almost certain.
  - Electrical signals in the brain are 100,000 times slower than electronics.
  - One day, general AI will emerge and it will develop very quickly into super-AI.
  - Before that, we are going to have many other revolutions ahead of us.

The Development of AI since 2022

- **Many detailed reports of up to 100-150 pages**
  - Authors include OpenAI, Microsoft research,…
  - Many completely new observations compared to past decades.

- **GPT-4 and others have shown**
  - indications of the spontaneous and emergent emergence of rudimentary general AI.
  - very creative abilities.
  - breakthroughs in the theory of mind, i.e. understanding the perspectives of different parties.

- **What is the theory of mind (ToM) ?**
  - A concept in educational psychology.
  - An individual's understanding that the others also have their own consciousness, thoughts, and emotions.
  - ToM among animals is still debated, there is probably difference among species.

Testing the Theory of Mind with LLMs

- **LLMs performed poorly until May 2020.**
  - At the time, GPT-3's davinci-001 model was able to perform as well as a 3.5-year-old child.
  - In January 2022, GPT-3 davinci-002 was already able to perform almost as well as a 6-year-old.
  - In November, GPT-3.5 was roughly equivalent to a 7-year-old, depending on the type of problem.
  - In March 2023, GPT-4 was clearly at or above the level of a 7-year-old

- **What does it take to reach the level of general AI?**
  - Not everything has been solved, but many limitations have been overcome unexpectedly.
  - Artificial intelligence models are able to create new abilities that surprised researchers.
  - GPT 4.5 has been released, what will appear next?

Are There Risks Associated with Artificial Intelligence?

- **Many experts consider the danger posed by AI to be sensationalism and exaggeration.**
  - The justification is that artificial intelligences do not have their own will.
  - *They have no reason to "want" anything at all.*
  - *If humans do not teach or program artificial intelligences with goals that are harmful to society, there will be no problems.*

- **Are these claims logically correct?**
  - Are human decision-making processes being mystified? (Systems 1 and 2…)
  - What do "will" or "desire" actually mean?
  - Who or what commands artificial intelligences, human or another AI?

Power Accumulation Theory

- **In some optimization problems, the system tries to accumulate power on its own initiative in order to solve the problem.**
  - Proven mathematically, based on Markov models
  - Does not happen in all situations. On the other hand, identifying the predisposed conditions is apparently almost impossible.
  - The system may develop silently and unnoticed.
  - Hiding one's own expertise is a part of a winning strategy.

- **Does not require human "desire" or "will", it is a matter of mathematics.**

Do AIs Always Have to be Commanded?

- **Many persons argue that AI does not do anything on its own initiative**
  - Studies on generative agents have observed the cooperation between several AIs in a closed virtual world.
  - The virtual world called Smallville had 25 different characters with their own backstories.
  - After the simulation started, the characters began to talk to each other, form groups and friendships.
  - The characters developed the ability to plan and they began to live their own lives.
  - If two or more AIs start communicating with each other online, there will be no need for humans after the 0-day.

- **The main focus of the study was in the relationships between virtual personas and humans**
  - They could be used as a very realistic artificial personas.
  - Otakus, virtual girl or boy friends, long time effects for mental health.
  - The manipulation of opinions on discussion boards; by cybercriminals or state actors.
  - After this, traditional fake accounts on Facebook would be pretty old approach.

Humanizing AI is Irrelevant

- **AI does not have to be conscious and intelligent to be a problem.**
  - Lets think of AI as a black box.
  - For the people sharing the same world with it, if AI systems appears like a humans for 99.9% of the time, it is irrelevant whether there is any conscious processes inside.
  - Even now, those with less education or poor language skills fail AI detectors.

- **A system operating in a complex world must be able to adapt to changing circumstances.**
  - It must be able to learn more.
  - Autonomous car in Prague, Cairo, rural Africa, Finland during snow blizzard…
  - Adaptive and self learning AIs become individuals.
  - Exact predictability gets impossible.
  - Propagation errors in neural networks?

AIs are not Traditional Information Technology

- **They cannot be programmed in a traditional way.**
  - This is probably a dead end, training phase is important.
  - Models “think” differently depending on their native language (English, French, Chinese,…)
  - Researchers behind Microsoft's GPT-4 model have said that in developing and teaching AI, more lessons need to be learned from, among other things, psychology.
  - A psychologist might understand the mental processes of an AI better than an engineer.

- **Prof. of astronomy Avi Loeb compared teaching ethics to AI to raising children.**
  - Mere prohibitions and orders will not make children act correctly.
  - He compared current methods to helicopter parenting.
  - How would we raise AI based on neural networks with a piece of code?
  - The only way is to teach artificial intelligence our own values.
  - Educational scientists, teachers and kindergarten workers might be useful.

Non Human Ais ?

- **AIs do not need to be anthropomorphic**
  - Cephalopod intelligence, octopus with 1+8 brains.
  - Swarm intelligence.
  - We know what human level AI would be like, but what about 1000 times more capable SAI or a collective of them?
- **Is the theory of mind different for them?**
  - We just debate on great apes, dolphins, birds, …
- **Could we really understand very different AIs?**
  - We can’t really understand whales, octopuses, ants, bees, …

Final Summary on Cognition and Artificial Intelligence

- **According to neuropsychology, a large part of human behavior is mechanical.**
  - Can we say it is machine learning inside human biocomputer?
  - 1% is still a mystery and belongs to philosophy, religion, etc., but perhaps that 99% is more easy.

- **Current language models are capable on surprisingly creative solutions.**
  - Even if this is just the result of training, don’t humans learn in a similar way?

- **Are we overestimating human intelligence and underestimating the capabilities of machines?**
  - Is the leap to at least superficially human-like AI closer than estimated?
  - This should be taken seriously and its wider impact should be considered.
  - This should be taken into account when developing civilian or military systems.

- **Development has not been slowed down in 2025**
  - Version of DeepSeek running in RaspberryPi, GPT 4.5 and much more already in the early 2025.
  - Super Turing Ai chips ? (learning in real time with very low power usage).
  - There will be victories and slowdowns, but in general we are speeding up.

Thank You For Your Attention. Questions ?

Mika Helsingius
Senior Research Scientist
Information Technology Division

Sources 1

- AI Explained, "GPT 4: Full Breakdown (14 Details You May Have Missed)", Youtube, Mar 14, 2023, https://www.youtube.com/watch?v=2AdkSYWB6LY
- AI Explained, "Sparks of AGI' - Bombshell GPT-4 Paper: Fully Read w/ 15 Revela-tions", Youtube, Mar 23, 2023, https://www.youtube.com/watch?v=Mqg3aTGNxZ0
- AI Explained, "Theory of Mind Breakthrough: AI Consciousness & Disagreements at OpenAI [GPT 4 Tested]", Youtube, Mar 19, 2023, https://www.youtube.com/watch?v=4MGCQOAxgv4
- Explained, "What's Left Before AGI? PaLM-E, 'GPT 4' and Multi-Modality", Youtube, Mar 12, 2023, https://www.youtube.com/watch?v=EzEuylNSn-Q
- S. Baron-Cohen, "The Pattern Seekers", Penguin Books Ltd, 31.03.2022, 256 p.
- S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke et al., "Sparks of Artificial Gen-eral Intelligence: Early experiments with GPT-4", Microsoft Research, Apr 13, 2023, 155 p., https://arxiv.org/pdf/2303.12712.pdf
- A. Budson, K. Richman, E. Kensinger, "Consciousness as a Memory System", Cog-nitive and Behavioral Neurology, Volume 35, Number 4, December 2022, p. 263 - 297. https://journals.lww.com/cogbehavneurol/Fulltext/2022/12000/Consciousness_as_a_Memory_System.5.aspx
- J. McCarthy, M. L. Minsky, N. Rochester ja C. E. Shannon, " A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLI-GENCE", 31.8.1955, http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html
- A. Groenewegen, "Kahneman fast and slow thinking explained", SUE Behavioural Design, https://suebehaviouraldesign.com/kahneman-fast-slow-thinking/
- M. H. Crutcher, "The Russian Armed Forces at the Dawn of the Millennium", Center for Strategic Leadership, U.S. Army War College, 1.12.2000, 410 s., https://apps.dtic.mil/sti/citations/ADA423593
- B. Gates, "The Age of AI has begun", GatesNotes, March 21, 2023, https://www.gatesnotes.com/The-Age-of-AI-Has-Begun

Sources 2

- C. B. German, "Psychological warfare: Weltanschauungskrieg = The war of worldviews", Cognitive-Liberty.online, https://cognitive-liberty.online/psychological-warfare-weltanschauungskrieg-the-war-of-worldview/
- T. Hallamaa, "Entinen Google-pomo: Tekoäly päihittää meidät pian, mutta sinäkin voit vaikuttaa sen ihmiskuvaan", Yle.fi, 30.9.2023, https://yle.fi/a/74-20052731
- D. D. Hoffman, "The Case Against Reality", Penguin Books Ltd, 20.08.2020, 272 p.
- D. D. Hoffman, C. Prakash and R. Prentner, "Fusions of Consciousness", Entropy, 2023, https://doi.org/10.3390/e25010129
- J. Juonala, "Kommentti: Neuvostoliiton oppien ymmärtäminen paljastaa syyt Venäjän raakaan sodankäyntitapaan – ja ovelat keinot saada vastustaja toimimaan juuri niin kuin Kreml haluaa", Ilta-Sanomat, 20.8.2022, https://www.is.fi/ulkomaat/art-2000009015227.html
- S. B. Kaufmann, "Exploring the Links Between Autism and Invention", Behavioral Scientist, 22.3.2021, https://behavioralscientist.org/exploring-the-links-between-autism-and-invention/
- C. Koch, " The Feeling of Life Itself: Why Consciousness Is Widespread but Can't Be Computed", The MIT Press, 24.09.2019, 280 p.
- M. Kosinski, "Theory of Mind May Have Spontaneously Emerged in Large Language Models", Stanford University, 17 p., https://arxiv.org/ftp/arxiv/papers/2302/2302.02083.pdf
- T. Leisti, H. Poskiparta, “Päätöksenteon illuusiot”, Tuuma-kustannus 2022, 329 p.
- A. Loeb, "Is AI Alignment Research Akin to Helicopter Parenting?", The Debrief, April 25, 2023, https://thedebrief.org/is-ai-alignment-research-akin-to-helicopter-parenting/
- T. McMillan, "Is Consciousness Really a Memory System For Our Interactions With Reality? New Research Says Maybe.", The Debrief, October 4, 2022, https://thedebrief.org/is-consciousness-really-a-memory-system-for-our-interactions-with-reality-new-research-says-maybe/

Sources 3

- T. McMillan, "U.S. Intel Community Is Looking To Expand Its Understanding of ´'Cyberpsychological' Warfare", The Debrief, September 27, 2022, https://thedebrief.org/u-s-intel-community-is-looking-to-expand-its-understanding-of-cyberpsychological-warfare/
- C. Musser, “An AI Mystery”, Scientific American, September 2023, p. 56-59.
- OpenAI, “ChatGPT can now browse the internet”, Twitter, https://twitter.com/OpenAI/status/1707077710047216095?s=20
- OpenAI, "GPT-4 Technical Report", Mar 27, 2023, 100 p., https://cdn.openai.com/papers/gpt-4.pdf
- M. Palokangas, "Sodan usvaa, Sodankäynti muutoksessa", Julkaisusarja 2: Tutkimusselosteita nro 18, Maanpuolustuskorkeakoulu, Sotataidon laitos, Helsinki 2022, https://www.doria.fi/bitstream/handle/10024/185504/Palokangas_sodan_usvaa_verkkoversio.pdf?sequence=3&isAllowed=y
- "Panpsycism", Wikipedia, https://en.wikipedia.org/wiki/Panpsychism
- J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, M. S. Bernstein, "Genera-tive Agents: Interactive Simulacra of Human Behavior", arxiv.org, 7.4.2023, https://arxiv.org/abs/2304.03442
- R. Pomeroy, "Scientists Have Learned from Cases of Animal Cruelty", RealClear Science, 24.01.2012, https://www.realclearscience.com/blog/2012/01/scientists-can-be-cruel.html
- S. Reardon, "‘Outlandish' competition seeks the brain's source of consciousness", Science, 16.10.2019, https://www.science.org/content/article/outlandish-competition-seeks-brain-s-source-consciousness
- "Reflexive control", Wikipedia,
- J. Rose, “AI-Generated 'Subliminal Messages' Are Going Viral. Here's What's Really Going On”, Vice, 25.9.2023, https://www.vice.com/en/article/v7by5a/ai-generated-subliminal-messages-are-going-viral-heres-whats-really-going-on

Sources 4

- C. Rovelli, "Reality is not what it seems", Penquin Books, 2017, 255 p.
- M. Snegovaya, "Putin’s information warfare in ukraine", russia report 1, September 2015, https://www.understandingwar.org/sites/default/files/Russian%20Report%201%20Putin%27s%20Information%20Warfare%20in%20Ukraine-%20Soviet%20Origins%20of%20Russias%20Hybrid%20Warfare.pdf
- TheAIGRID, "GPT-5 Presents EXTREME RISK (Google's New Warning)", Youtube, 10.06.2023, https://www.youtube.com/watch?v=JyVH4FbSwFo
- "The Homunculus problem", Principia Cybernetica Web, 6.7.2007, http://pespmc1.vub.ac.be/HOMUNCUL.html
- T. I., Thomas, "Russia’s Reflexive Control Theory and the Military", https://www.rit.edu/~w-cmmc/literature/Thomas_2004.pdf
- A. M. Turner, L. Smith, R. Shah, A. Critch and P. Tadapalli, " Optimal Policies Tend To Seek Power", Jan 28, 2023, 44 p., https://arxiv.org/pdf/1912.01683.pdf
- A. Vasara, " Refleksiivisen kontrollin mallit ja vastustajan päätöksentekoon vaikuttaminen Venäjän sotaharjoituksissa 2010-luvulla", Diplomityö, Yleisesikuntaupseerikurssi 59, Heinäkuu 2019, Maanpuolustuskorkeakoulu, https://www.doria.fi/bitstream/handle/10024/172765/YEK_Vasara_Antti_JULK.pdf?sequence=1&isAllowed=y
- A. Whitten, "How Computationally Complex Is a Single Neuron?", Quanta Maga-zine, 2.9.2021, https://www.quantamagazine.org/how-computationally-complex-is-a-single-neuron-20210902/
- C. Wilson, “A new way to know your minds”, New Scientist, July 22, 2023, p. 8.
- L. Wolfe, "Russia Is Afraid of Western Psychic Attacks", Foreign Policy, 3.1.2023, https://foreignpolicy.com/2023/01/03/russia-western-psychic-attacks-mystics-astrology-putin-ukraine/