HCAIM@BME

Human Centered AI Masters
at the Budapest University of Technology and Economics

Kitti Mezei, Mihály Héder, Péter Antal | 06. 28. 2023

humancentered-ai.eu

hcaim human centred artificial intelligence masters

Contents

hcaim human centred artificial intelligence masters

01 HC Reg

- Motivation for regulation
- External factors
- Internal factors

02 AI and law

- Why do we regulate AI?
- Exploring the boundaries of AI and law

03 HCAI

- What is AI?
- Problems with AI
- HCAI

04 HCAI education

- HCAIM @ BME
- HCAI @ BME
- HCAI in adult education

Motivations for Human Centered Regulation

Machines in the service of society

- It has been understood for centuries now that the wealth of societies is connected to their level of **technological sophistication**
  - human (or animal) efficiency can be surpassed by machines
- Industrial revolutions: the rise of the **machine maker** scientist/engineer
- 21st century: the rise of the machine maker with **humanities** skills

Rijn en Zon, Utrecht (FREEPIK)

Internal pressures on Machine Makers
What is Human-Centered Design of machines?

What is not human-centered design?

Source: Elias Beck. *'Child Labor in the Industrial Revolution'*. History Crunch. December 30, 2021. https://www.historycrunch.com/child-labor-in-the-industrial-revolution.html#/

What is not human-centered design?

Gilbreth **chronocyclograph** of motions necessary to move and file sixteen boxes full of glass, n.d. From: Mike Mandel, *Making Good Time: Scientific Management, the Gilbreths, Photography and Motion, Futurism* (Santa Cruz, CA: California Museum of Photography, University of California, Riverside, 1989), 26.

What is the engineers role on all of this?

- Self-absolving strategy 1
  - “The imperative of technology” - This is what efficiency dictates
- Self-absolving strategy 2
  - This is what the customer wants
- Self-absolving strategy 3
  - It was legal at the time and by the way I don't know anything about law

- Human-Centered Design: **rejection of all the above**
  - **humanities toolkit**
    - **argumentation**
    - **critical thinking**

Beginnings of HCD

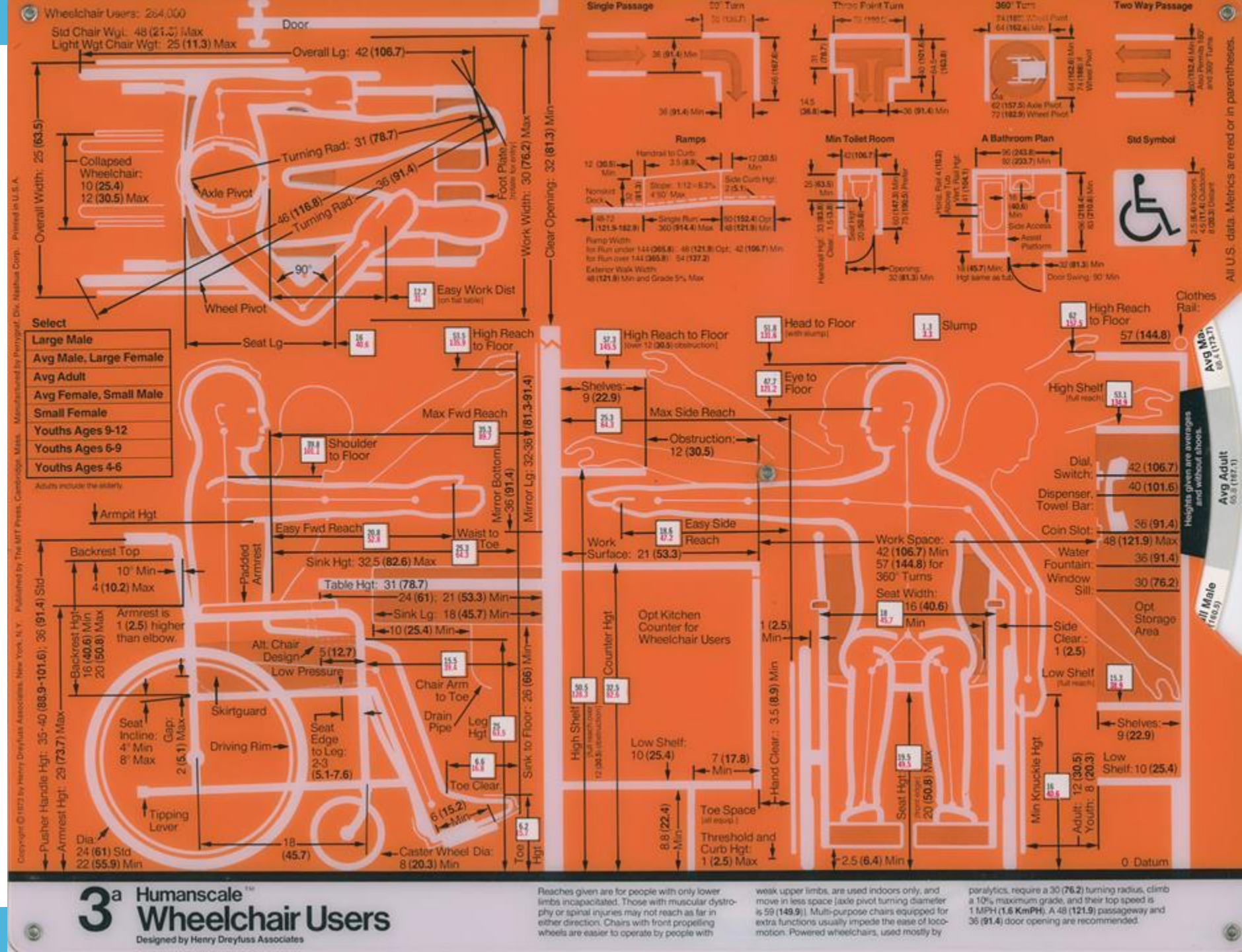


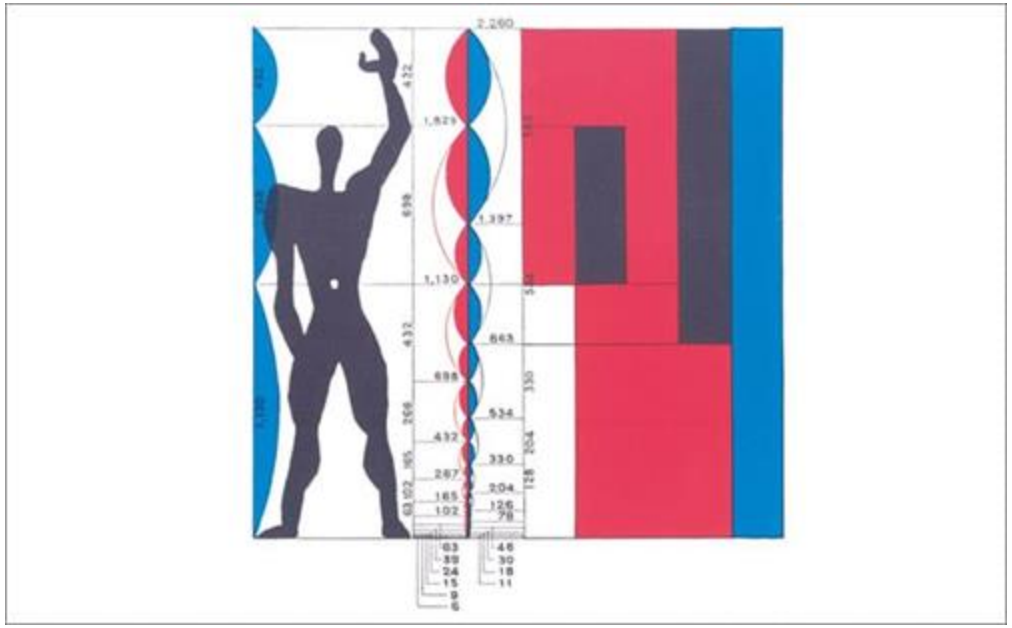
Image: Henry Dreyfuss Associates, *Humanscale* selector 3a “Wheelchair Users,” 1974. Plastic, paper, and metal. Milwaukee Art Museum Research Center.

Source: Hanna Pivo, 20th-Century Tools for Measuring Time and Bodies April 19, 2019 https://blog.mam.org/2019/04/19/20th-century-tools-for-measuring-time-and-bodies/

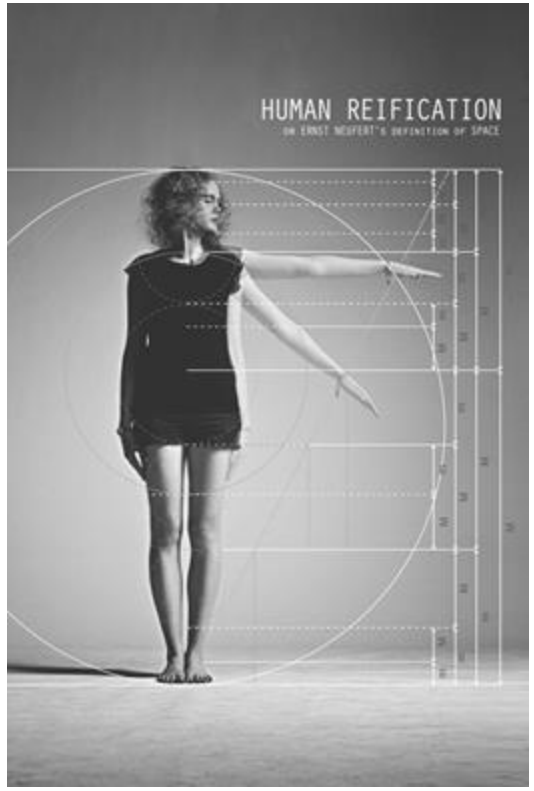
Why do we need to deal with AI specifically?

human centred artificial intelligence masters

- There already have been techno-ethical questions and **human-centered design**
  - **Ergonomy**
  - **UX design**



The Modulor by Le Corbusier 1943-54

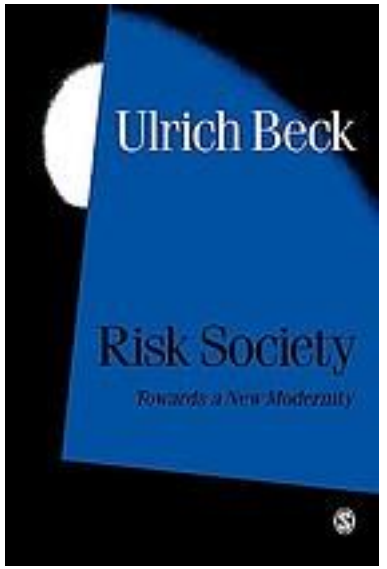


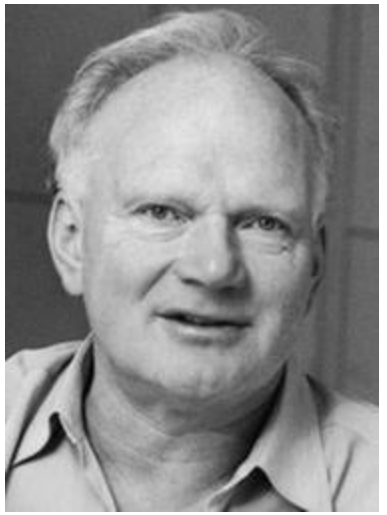
Human Reification - Paul Gisbrecht

External pressures on Machine Makers Modernity 2.0

Ulrich Beck's analysis (1980's)

- A main feature of modern society is that it is **preoccupied with the future**, and
  - especially the negative scenarios, that is 'risks'
- Catastrophes were formerly attributed to bad luck or divine acts
  - but not in Humanity's control
- Now that our control seems greater (modern science) the **responsibility is ours**
  - this in turn undermines the institutions of modern society, e.g. trust in science





1944-2015

Modernity 2.0

- The victories of the first modernity (taking risks) have a boomerang effect
- **Taking risks** not serve us anymore
- So we enter **reflexive modernity**
  - Pesticide
  - Ozone
  - Nuclear
  - Toxins
  - CFC
  - Plastics, etc.
- Plus, we anticipate even more negative consequences
  - AI, GMO

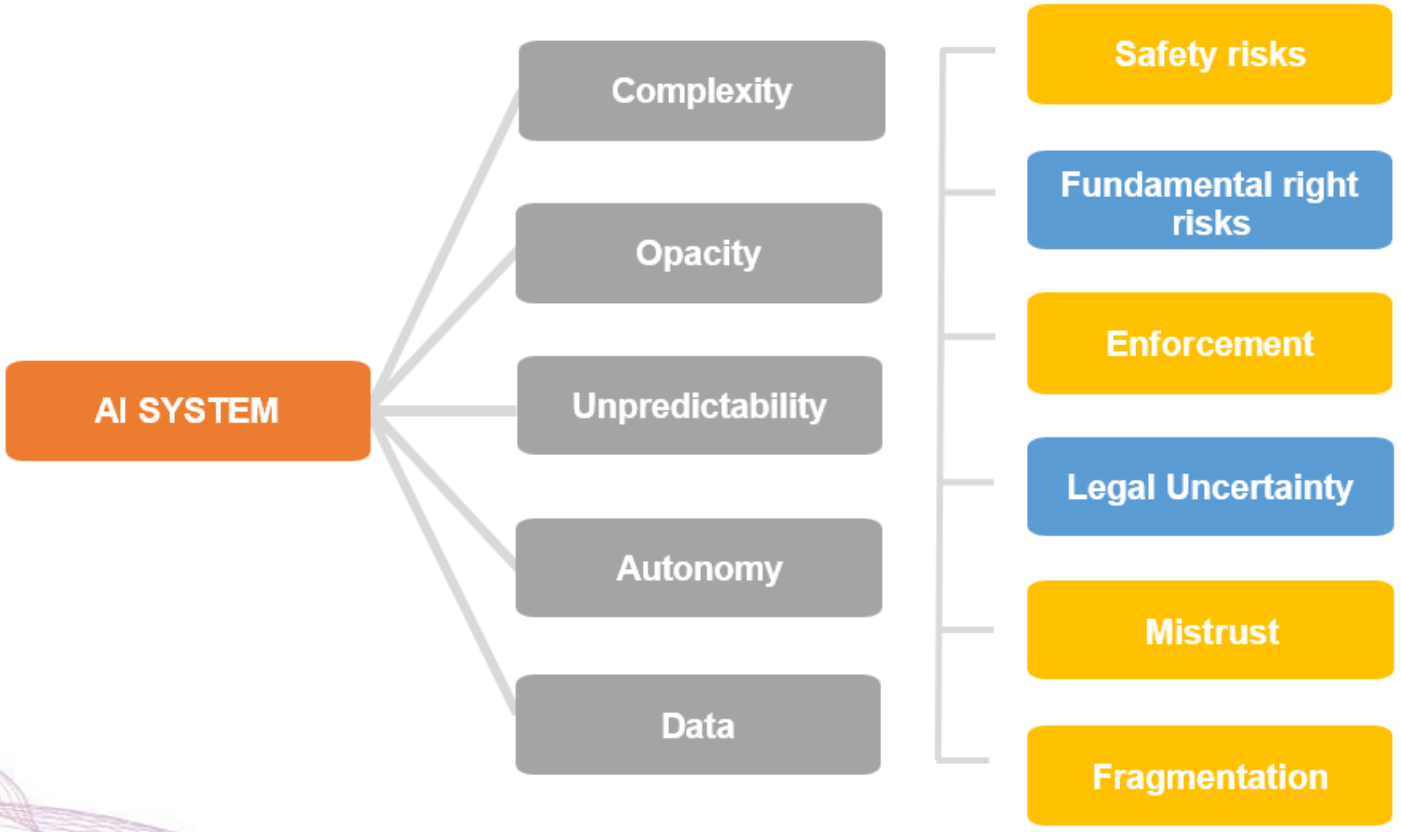
Modernity 2.0

- No clear culprits
  - many of us are implicated in these negative effects

  we need modern technology to even identify and tackle risks
- The new risks are far more evenly distributed
  - Modernity 1
    - Living next to a factory was risky, if rich you could move away
  - Modernity 2
    - Ozone, global warming
    - arguably you can only temporarily can avoid these risks with your personal wealth

AI and law

Why do we regulate AI use cas



human centred artificial intelligence masters

European Commission

Exploring the Boundaries of AI and Law

human centred artificial intelligence masters

Are we embracing the wild west or building a digital utopia?

How can regulatory approaches for technology strike the perfect balance between fostering innovation and ensuring responsible development?

Proactive collaboration

Agile and adaptive frameworks (future proof)

Risk-based assessments

Ethical and human-centered focus

Encouraging innovation through sandboxes

Global cooperation and standards

Exploring the Boundaries of AI and Law

Why is ethics a key issue for the AI industry (ethics by design)?

How can ethical considerations be effectively incorporated into AI regulation frameworks to ensure fairness, transparency, and accountability in AI systems?

What responsibilities do organizations have to ensure fairness and non-discrimination in AI?

Ethical guidelines and principles, ethical assessments

Requirements for explainability and interpretability of AI systems

Adherence to non-discrimination principles (anti-discrimination law), discriminatory practices

Regular auditing of algorithms for biases (by who and when? by an internal and/or external auditor?)

The establishment of clear lines of responsibility and accountability for AI system outcomes

Exploring the Boundaries of AI and Law

human centred artificial intelligence masters

What mechanisms and guidelines should be established to address the issue of accountability and liability in cases involving AI, particularly in complex, dynamic environments?

Clear legal frameworks that define the responsibilities and liabilities of different stakeholders in AI development, deployment, and operation

Regulatory oversight

Transparency

Continuous monitoring and auditing

Adequate insurance coverage

Exploring the Boundaries of AI and Law

human centred artificial intelligence masters

Can responsibility for loss or damage caused by AI be attributed to someone? What are the potential civil law or criminal law liabilities?

Should there be specific liability frameworks for AI systems and their developers, manufacturers, or operators?

How does the use of AI technologies impact the existing intellectual property framework, and what are the implications for the protection, enforcement, and commercialization of intellectual property rights in the context of AI-driven innovations?

Exploring the Boundaries of AI and Law

How can individuals' privacy be protected when AI systems rely on vast amounts of personal data for training and operation? What measures should be in place to ensure compliance with data protection laws?

Who should be held accountable for data protection in the context of AI (data controllers, AI developers, or service providers)?

How can individuals be protected from potential adverse effects resulting from automated profiling and decision-making processes?

How can the security of data used by AI systems be ensured to prevent unauthorized access, data breaches, or misuse?

Human-centered AI engineering

hcaim human centred artificial intelligence masters

What is AI? (~1950-2010)

human centred artificial intelligence masters

Standard model

- Humans are intelligent to the extent that our actions can be expected to achieve our objectives
- Machines are intelligent to the extent that their actions can be expected to achieve their objectives [S.Russell, 2018: AI25]

Universal Turing machine

- Finite operations over a finite alphabet
- Universality
- Incompleteness (truth≠provability)
- Space and time complexity classes

(Bayesian) Decision theory

- Probability theory
- Utility theory
- Optimal decision
- Bayesian inference/model averaging

Turing test

- 'Can machines think?'
- Turing, Alan (October 1950), "Computing Machinery and Intelligence", Mind, LIX (236): 433–460
- Imitation game (~chat)

Problems with AI: 2010<

human centred artificial intelligence masters

Technological

- Symbolic vs. Subsymbolic paradigm
- White vs. Black-box models
- Associative vs. Causal approaches
- Flat vs. Hierarchical
- Narrow vs. General intelligence

Developmental

- Understandability
- Compositionality
- Workflow
- Group work

Legal

- Safety
- Provably beneficial
- Fairness and bias
- Responsibility

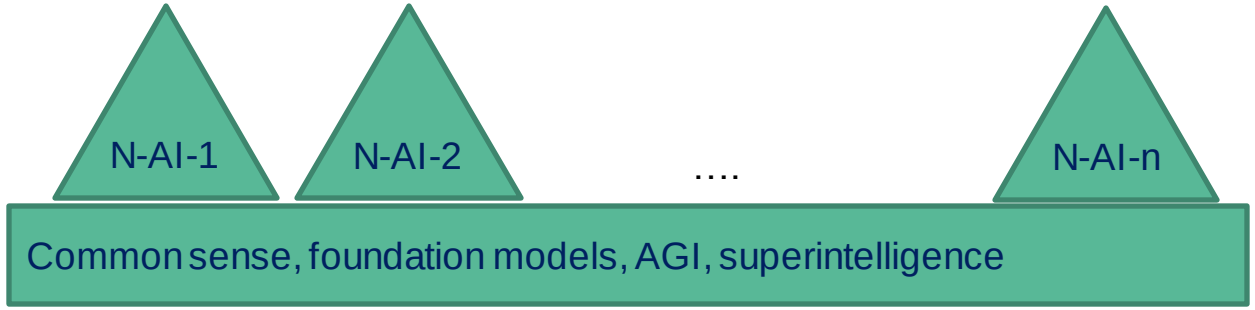
Societal

- Digital addiction
- Polarization of the society
- Fake news
- Subliminal effects
- Advertisement-driven attention economy
- Superintelligence/artificial general intelligence

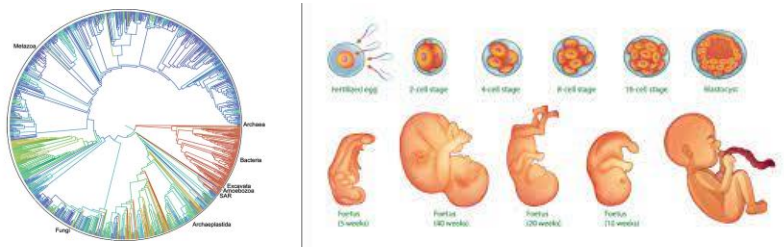
Artificial general intelligence

Artificial general intelligence (AGI): it can learn to accomplish any intellectual task that human beings or animals can perform.

- Vision, robotics, natural language processing (NLP)
- Self-driving cars, automation of scientific discovery,...



Domain-specific General Problem Solvers (GPS, 1957)

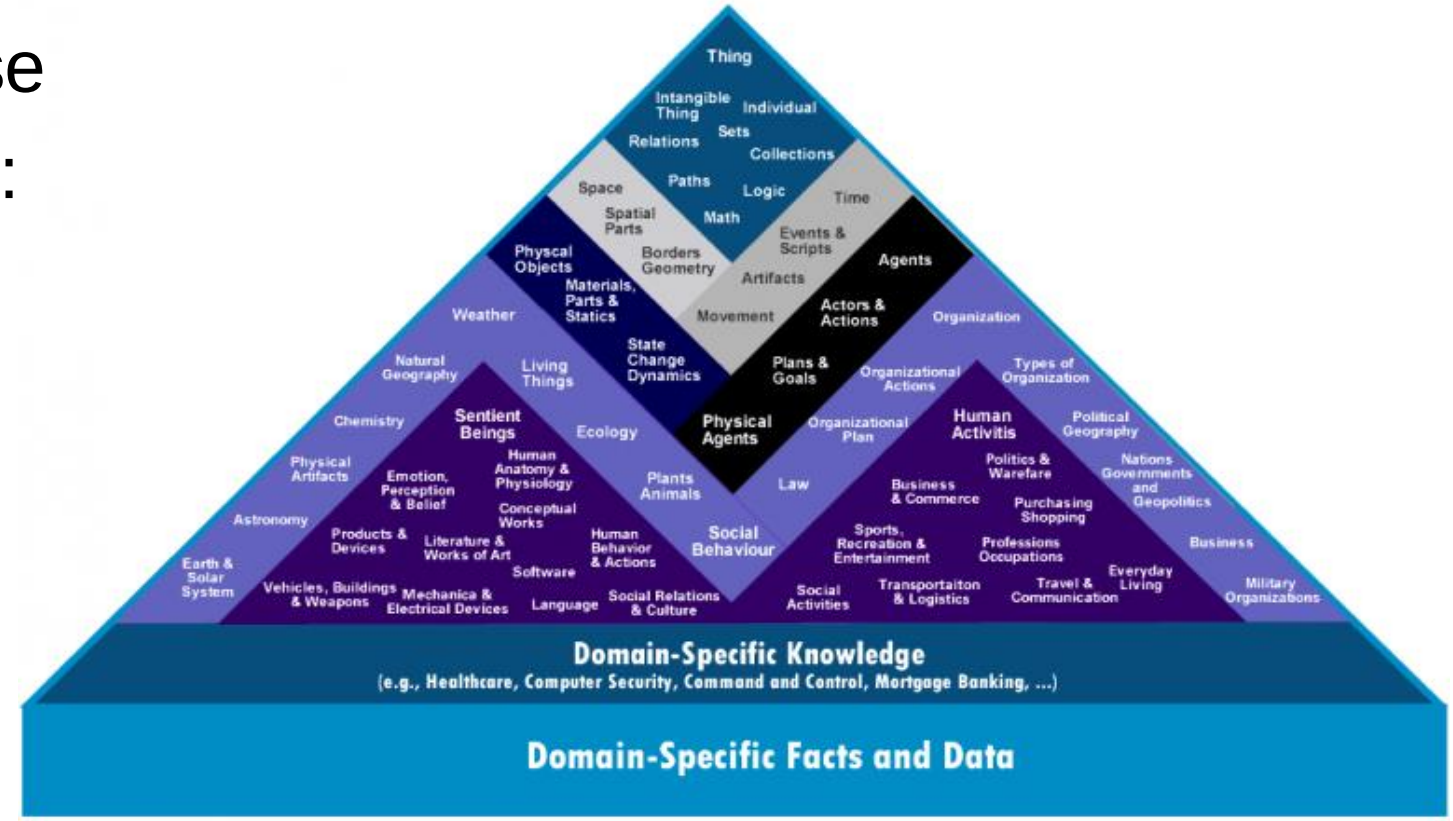


Current biomass (C): $\times 10^{11}$ tonnes
#DNA base pairs: 10^{37}
#species: ~10^{7} (extinct: x10^{9})
#animals: ~10^{19}
#cells in human body: ~10^{14}
#neurons in brain: ~10^{11}
#synapsis per neuron: ~10^{4}

Knowledge bases: Encyclopedists (~1750), WorldBrain (1936), Naive physics manifesto(1979), CYC(1984-), Wikimedia(2003), GitHub(2007)
AGI systems: IBM Watson(2011), WolframAlpha(2009), AlphaZero(2017), chatGPT(2018-),...

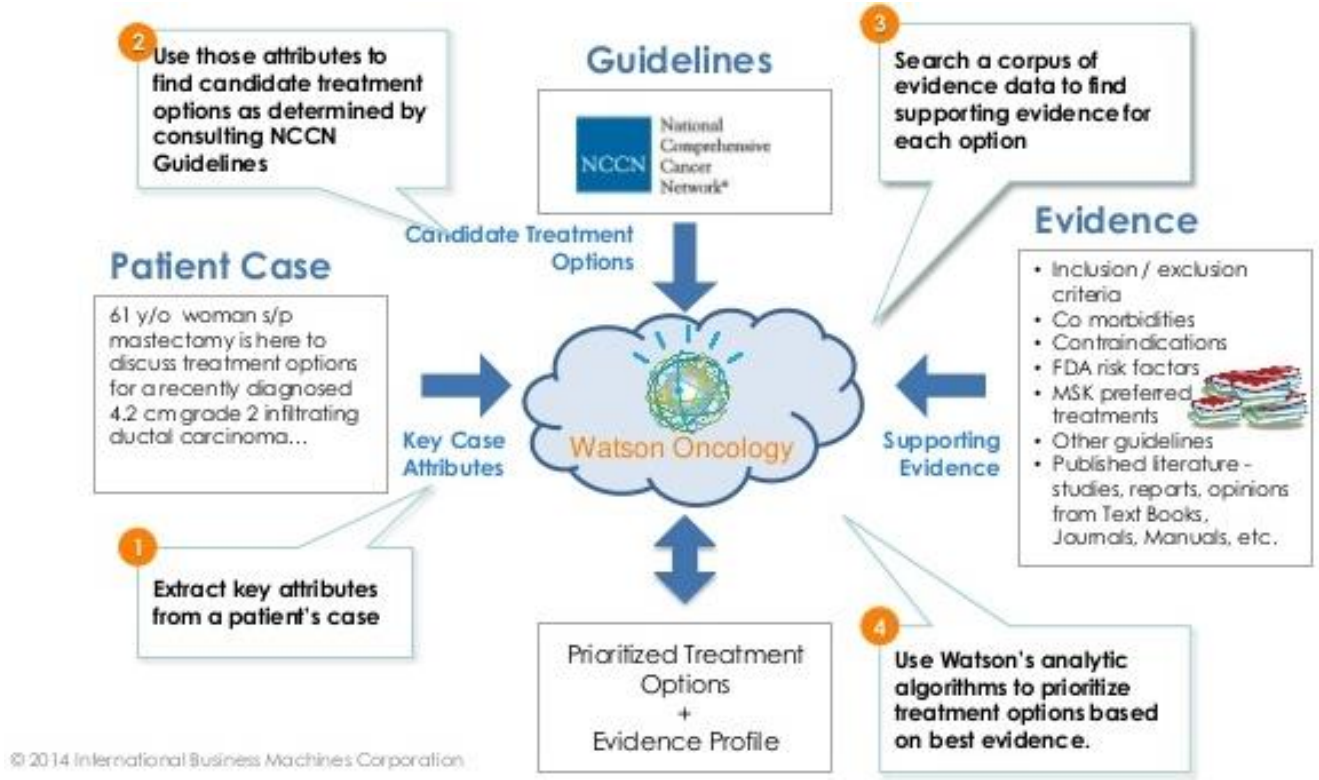
The Cyc project (1984-2016)

- Goal: common sense
- Estimations in 1984:
  - 250 000 rules
  - 350 man-year
- Language: CycL
- Access: OpenCyc
- State (~2020)
  - 239,000 concept
  - 2,093,000 facts

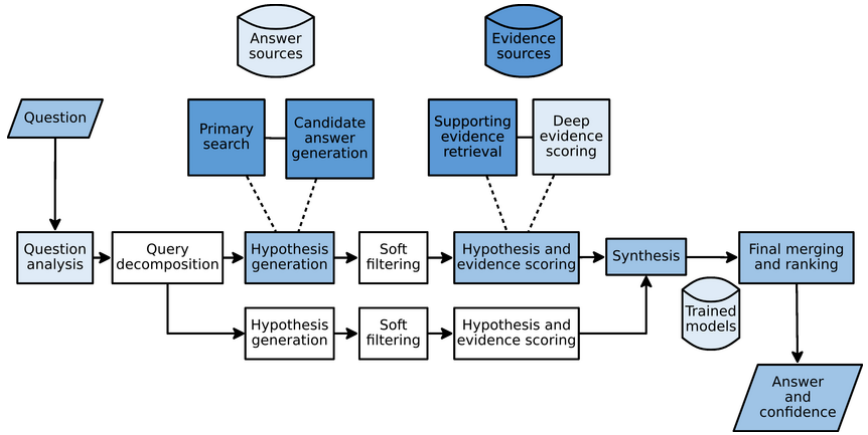


Lenat, Douglas. "Creating a 30-million-rule system: Mcc and cycorp." *IEEE Annals of the History of Computing* 44.1 (2022): 44-56.

IBM Watson (2011)



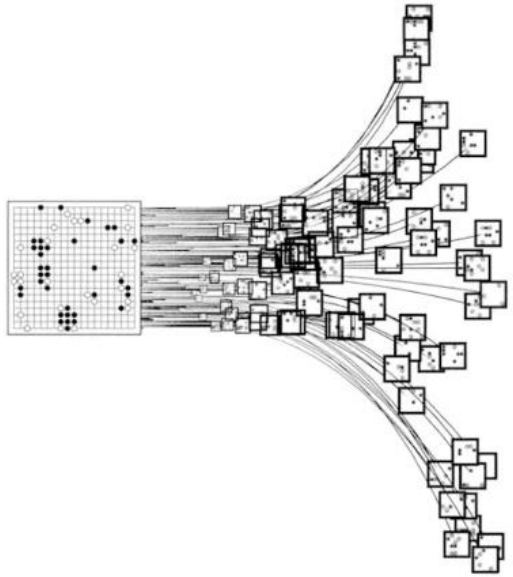
- Natural language processing
- Inference
- Game theory



Strickland, E. (2019). IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*, *56*(04), 24-31.

AlphaGo (2017)

- Google DeepMind
- Monte Carlo tree search
- 2016: 9 dan
- 2017: wins against human champion



Silver, David, et al. "Mastering the game of Go with deep neural networks and tree search." *nature* 529.7587 (2016): 484-489.
Silver, David, et al. "Mastering the game of go without human knowledge." nature 550.7676 (2017): 354-359.

Generative pre-trained transformers

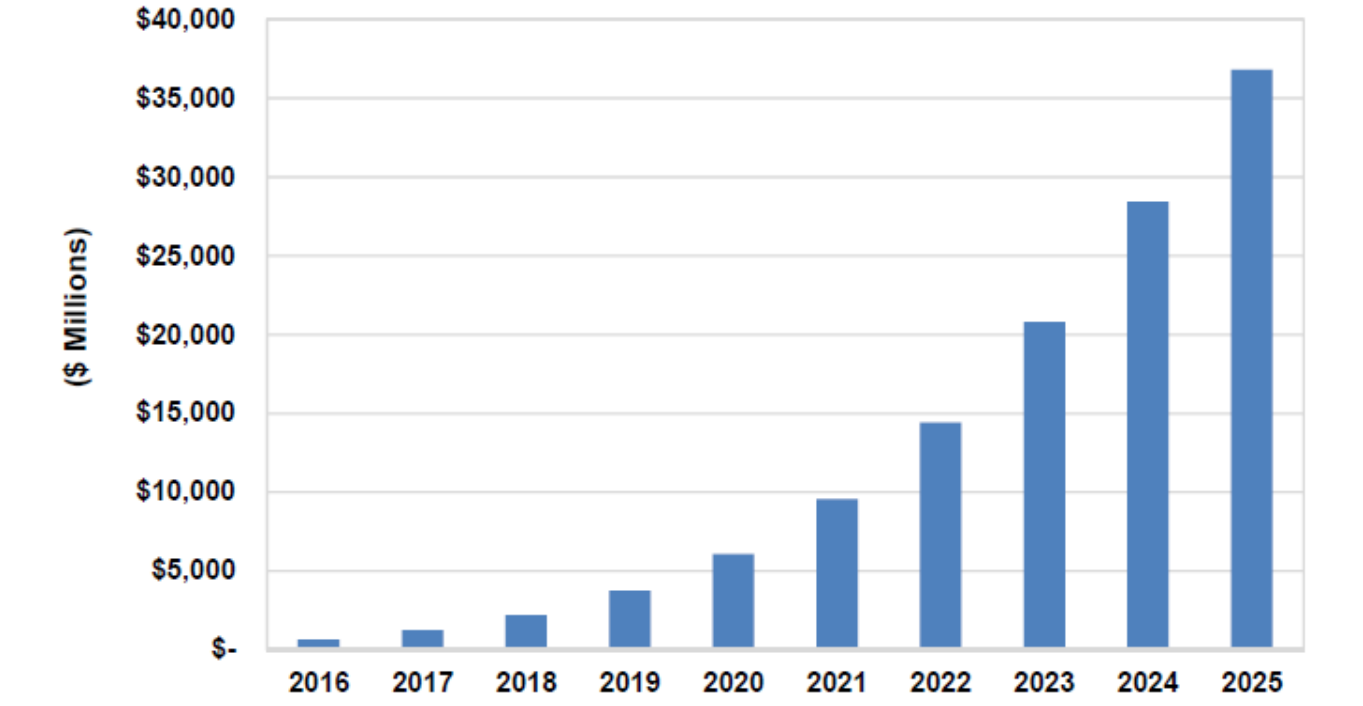
- Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).
- Radford, Alec, et al. "Improving language understanding by generative pre-training." (2018).
- Subramanian, Sandeep, et al. "**Learning general purpose distributed sentence representations via large scale multi-task learning**." arXiv preprint arXiv:1804.00079 (2018).
- Brown, Tom, et al. "**Language models are few-shot learners.**" Advances in neural information processing systems 33 (2020): 1877-1901.
- Ouyang, Long, et al. "**Training language models to follow instructions with human feedback**." Advances in Neural Information Processing Systems 35 (2022): 27730-27744.
- Bubeck, Sébastien, et al. "**Sparks of artificial general intelligence: Early experiments with gpt-4.**" arXiv preprint arXiv:2303.12712 (2023).
- Luo, Renqian, et al. "BioGPT: generative pre-trained transformer for biomedical text generation and mining." Briefings in Bioinformatics 23.6 (2022): bbac409.

Financial resources

human centred artificial intelligence masters

1-10 bn$ monthly investment in AGI (2023)!

Chart 1.1 Artificial Intelligence Revenue, World Markets: 2016-2025



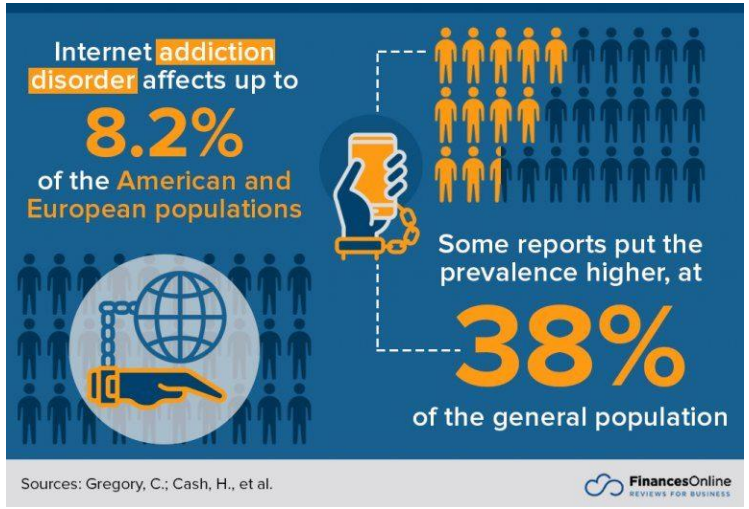
(Source: Tractica)

Digital addiction

human centred artificial intelligence masters

No need for AGI to ruin humanity/democracy/mental health; basic AI is enough: addiction, bias, fairness, privacy, polarization, fake news,...

Reinforcement learning in action:
training brains in the attention economy



https://financesonline.com/technology-addiction-statistics/

Problems with electricity (~1900)

human centred artificial intelligence masters

Technological

- Generation
- Transformation
- Transfer
- Storage
- Usage

Developmental

- Maintainance
- Institutional

Legal

- Safety
- Responsibility
- Universal access (at minimum level)

Societal

- Health
- Electric shock
- Electromagnetic effects
- Urban environment
- Economy

Problems with drugs (~1900)

human centred artificial intelligence masters

Technological

- Discovery
- Synthesis
- Formulation
- Storage
- Usage

Developmental

- Mechanism of action
- Target proteins/pathways
- Distillation/extraction
- Synthesis

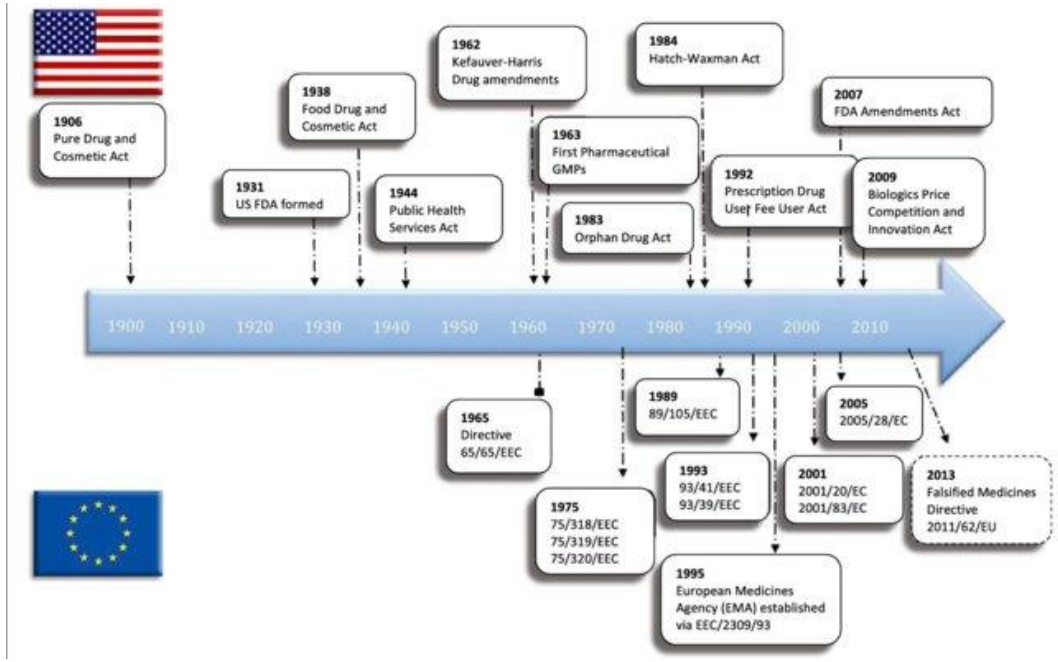
Legal

- Safety
- Responsibility
- Provably beneficial
- Pricing
- Universal access (at minimum level)

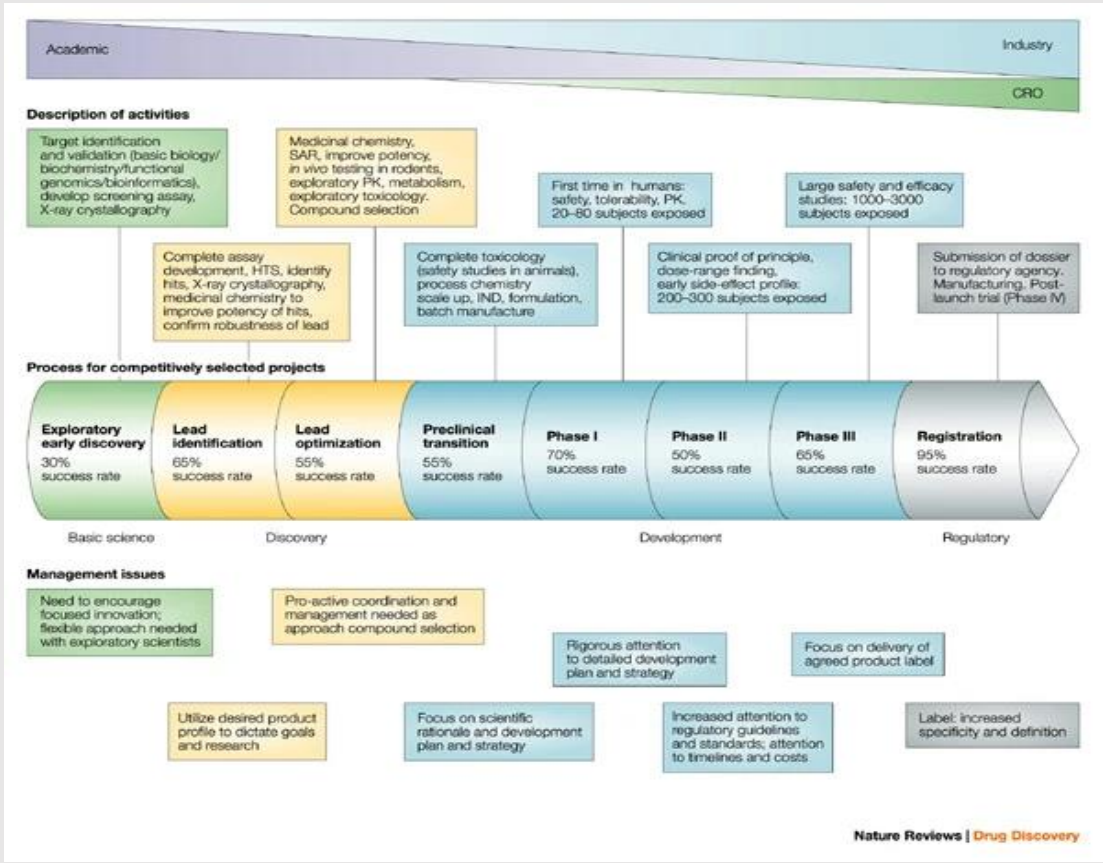
Societal

- Disease burden
- Side-effects
- Addiction
- Miracle cures

Regulation for drug discovery



Dunne, Suzanne, et al. "A review of the differences and similarities between generic drugs and their originator counterparts, including economic benefits associated with usage of generic medicines, using Ireland as a case study." *BMC Pharmacology and Toxicology* 14 (2013): 1-19.



Nwaka, Solomon, and Robert G. Ridley. "Virtual drug discovery and development for neglected diseases through public–private partnerships." *Nature Reviews Drug Discovery* 2.11 (2003): 919-928.

HCAI interpretations

hcaim human centred artificial intelligence masters

Human-computer interaction

- Man-machine interfaces
- Man-machine hybrids
- Linguistic interfaces
- Sensorimotoric/brain-computer interfaces
- Augmented reality

Human-compatible AI

- Intelligence explosion
- Artificial general intelligence (AGI)
- Superintelligence
- Existential risk
- Value alignment
- Provably beneficial AI

Human-centered AI (HCAIM)

- **Human rights (mental health,..)**
- **Democratic society (fake news,...)**
- **Trustworthy/Explainable AI**
- **Human-computer cooperation**
- **Collaborative workflows**
- **Auditing/approval (AI safety)**

Intelligence everywhere

- Smart wearable electronics
- Smart homes/cities
- Autonomous vehicles

Novel elements in HCAI

human centred artificial intelligence masters

Theory

- Federated learning, privacy-preserving learning
- Probabilistic programming, causal inference
- Machine teaching, collaborative inverse reinforcement learning, reinforcement learning with human feedback
- Multitask learning, transfer learning, foundation models, artificial general intelligence

Practice

- Automated programming
- Collaborative workflow systems
- Testing/Auditing

Ethics

- Rights for digital assistants/twins

Society

- The EU AI Act

HCAIM @ BME (2022<)

human centred artificial intelligence masters

The 60-credit EU-level HCAIM program is embedded into a 120-credit, 2-years M.Sc. program.

Course type			Course title	Neptun-cod	ECTS	Spec.				C	Vál.		
HCAIM	BME	BME				1	2	3	4			Min.	Max.
I.													
Basic	Common	From each group, the min. ECTS must be completed; From each group, up to the max. ECTS can be taken into account in the HCAIM 60 credits.	Applied algebra and mathematical logic	TE90MX75	5	5	5	5	5			5	5
			Mathematical Statistics	VISZMA11	5								
			Stochastics	TE90MX77	5								
	Specialization		Machine learning	VIMIMA27	5	5						5	5
			Machine learning		5						5		
			Deep learning	VITMMA19	5	5						5	15
			Application of Deep Learning in visual computing	VIIIMB10	5			5					
			Neural networks		5						5		
			Deep Learning in practice based on Python and LUA	VITMAV45	5						5		
			The security of machine learning	VIHIMB09	5					5		5	10
			Trusted artificial intelligence and data analytics	VIMIMB10	5					5			
	Elective		The ethics of artificial intelligence	GT41V105	2						2	2	2
			Artificial intelligence and the law	GT55V106	2						2	2	2
			Artificial General Intelligence	VIMIAV22	2						2	2	2
	Common		Project lab 2 (with AI content)		5	5	5	5	5			5	5
			Thesis work 1 (with AI content)		10	10	10	10	10				
			Thesis work 2 (with HCAI content)		20	20	20	20	20			15	15
			A. HCAIM basic, total		96	50	40	45	40	10	21	46	61
II.													
Opcio-nal	Common	Mandatory completion depending on specialization	Project lab 1 (with AI content)		5	5	5	5	5			0	5
	Specialization		Intelligent data analysis and decision support	VIMIMB09	5	5						0	5
			Business intelligence	VIAUMA24			5						
			AI-based human-machine interaction	VITMMA23					5				
			Machine learning case studies	VITMMA18	5	5						0	5
			Business intelligence lab	VIAUMB09			5						
			UX laboratory	VITMMB14					5				
			Advanced data analysis methods lab	VITMMB10	5	5						0	5
			B. HCAIM optional, total		20	20	15	5	15	0	0	0	20
			HCAIM basic + specialization optional, total									46	81
III.													
Opcio-nal	Elective	Recognisable											
			HCAIM optional, elective courses to the minimum 60 ECTS									14	0
Spec.	1	MIT-TMIT	Data science and artificial intelligence		Major								
	2	AUT	Software Development										
	3	IIT	Visual informatics										
	4	TMIT	User experience - UX and interaction		Minor								

HCAIM @ BME

hcaim human centred artificial intelligence masters

Participants

- Graduation at 2022.: 3 students
- Graduation at 2023 spring: 5 students
- Participants in 2023 spring: 19+11+2 students in their 1st/2nd/r3rd semester

Diploma Supplement

HCAIM certificate: **The student completed the requisite learning outcomes of the Human-Centred Artificial Intelligence Master's (HCAIM) programme, defined by the INEA/CEF/ICT/A2020/2267304 EU project.**

HCAIM@BME: Related programs

human centred artificial intelligence masters

Data science and AI M.Sc. specialization

Spring semester (mandatory)

- Machine learning 5 ECTS
- Intelligent data analysis and decision support 5 ECTS
- Advanced data analysis methods lab 5 ECTS

Fall semester (mandatory)

- Deep learning 5 ECTS
- Machine learning case studies 5 ECTS

Common (C) Elective

- Privacy and Security in machine learning 5 ECTS
- Trustworthy AI and data analytics 5 ECTS
- ...

Elective

- AI ethics 2 ECTS
- AI Law 2 ECTS
- Engineering Ethics 2 ECTS
- Artificial General Intelligence 2 ECTS
- ...

HCAIM@BME: Related programs

hcaim human centred artificial intelligence masters

Human-centred intelligent data analysis

A new Specialization in the EIT Digital Data Science Master School Programme

Fall semester (mandatory)

- AI Ethics 2 ECTS
- AI Law 2 ECTS
- Intelligent data analysis 5 ECTS
- Privacy and Security in machine learning 5 ECTS
- I&E Study 6 ECTS
- Thesis I 10 ECTS

Spring semester (mandatory)

- Trustworthy AI and data analytics 5 ECTS
- Thesis II 20 ECTS

Elective

- Engineering Ethics 2 ECTS
- Artificial General Intelligence 2 ECTS
- ...

eit Digital

Thanks!

Kitti Mezei, Mihály Héder, Péter Antal

Budapest University of Technology and Economics

hcaim human centred artificial intelligence masters