

Ethical and Social Aspects of AI

Piet Kommers
Martin Drlík
Eugenia Smyrnova-Trybulska
Vaida Masiulionytė-Dagienė

www.fitped.eu

2024



Erasmus+ FITPED-AI
Future IT Professionals Education in Artificial Intelligence
(Project 2021-1-SK01-KA220-HED-000032095)

Ethical and Social Aspects of AI

Published on

November 2024

Authors

Piet Kommers | Helix5, Netherland

Martin Drlík | Constantine the Philosopher University in Nitra, Slovakia

Eugenia Smyrnova-Trybulska | University of Silesia in Katowice, Poland

Vaida Masiulionytė-Dagienė | Vilnius University, Lithuania

Reviewers

Peter Švec | Teacher.sk, Slovakia

Oldřich Trenz | Mendel University in Brno, Czech Republic

Vladimiras Dolgopolas | Vilnius University, Lithuania

Erasmus+ FITPED-AI

Future IT Professionals Education in Artificial Intelligence

Project 2021-1-SK01-KA220-HED-000032095



**Funded by
the European Union**

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Slovak Academic Association for International Cooperation. Neither the European Union nor the granting authority can be held responsible for them.



Licence (licence type: Attribution-Non-commercial-No Derivative Works) and may be used by third parties as long as licensing conditions are observed. Any materials published under the terms of a CC Licence are clearly identified as such.

All trademarks and brand names mentioned in this publication and all trademarks and brand names mentioned that may be the intellectual property of third parties are unconditionally subject to the provisions contained within the relevant law governing trademarks and other related signs. The mere mention of a trademark or brand name does not imply that such a trademark or brand name is not protected by the rights of third parties.

© 2024 Constantine the Philosopher University in Nitra

ISBN 978-80-558-2223-5

TABLE OF CONTENTS

1 Introduction	7
1.1 Preface	8
1.2 Introduction	12
2 Historical Context	16
2.1 Historical context.....	17
2.2 Trust and Consciousness in AI.....	19
2.3 Rapid Course Rational	21
3 Industry 4.0.....	24
3.1 Industry 4.0.....	25
4 Citizens 2023	27
4.1 Citizens 2023	28
5 AI Needs a Blend of Disciplines and Values.....	30
5.1 AI needs a Blend of Disciplines and Values	31
6 Current State of the Art in OpenAI.....	35
6.1 Current State of the Art in OpenAI.....	36
6.2 ChatGPT: Optimizing Language Models for Dialogue.....	39
7 Adjacent Fields for Applying AI	40
7.1 Adjacent Fields for Applying AI	41
8 AI at the Crossroads in Human History	43
8.1 AI at the Crossroads in Human History	44
9 Landscape of AI Technologies	45
9.1 Landscape of AI Technologies	46
10 Machine Learning	48
10.1 Machine Learning	49
11 AI Taxonomies	52
11.1 AI taxonomies.....	53
12 Conversational Style of AI Systems.....	57
12.1 Conversational Style of AI Systems.....	58
13 AI in the Context of Human Sensing, Actions and Performance.....	60
13.1 AI in the Context of Human Sensing, Actions and Performance.....	61
14 AI's Impact on Human Employment	64
14.1 AI's Impact on Human Employment	65
15 Social Aspects of Artificial Intelligence	67
15.1 Social Aspects of Artificial Intelligence.....	68
16 Demystification of AI.....	70

16.1 Demystification of AI.....	71
17 The Relation Between AI and Empathy.....	73
17.1 The Relation between AI and Empathy.....	74
18 Empathy for Accountability of AI.....	77
18.1 Empathy for Accountability of AI.....	78
19 Towards a Virtual Friend.....	81
19.1 Towards a Virtual Friend.....	82
19.2 Concerns on the Relation Man versus Machine.....	85
20 The Bias of AI.....	87
20.1 The Bias of AI.....	88
21 How to Take Control of AI with Empathy and Transparency.....	90
21.1 How to take Control of AI with Empathy and Transparency.....	91
22 Empathy: Nothing Without Action.....	93
22.1 Empathy: Nothing without Action.....	94
23 The Target Goals of AI.....	96
23.1 The Target Goals of AI.....	97
24 Computational Thinking and AI.....	99
24.1 Computational Thinking and AI.....	100
25 The AI Cash-Flow Formula.....	102
25.1 The AI Cash-Flow Formula.....	103
26 With AI: No More Time for Stupid Tasks.....	105
26.1 With AI: No More Time for Stupid Tasks.....	106
27 Extra Attention to the Ethical Side.....	108
27.1 Extra Attention to the Ethical Side.....	109
28 AI Socratic Method.....	111
28.1 AI Socratic Method.....	112
29 Predicting when Invoices will be Paid: Payt's AI Solution.....	114
29.1 Payt's AI solution.....	115
30 AI in the Social Context.....	117
30.1 Artificial Intelligence in the Social Context.....	118
30.2 The Risk of Trivial Findings in Big Data.....	120
30.3 Current Applications of Artificial Intelligence.....	122
31 Definition of AI.....	124
31.1 Definition of AI.....	125
31.2 Singularity in Artificial Intelligence.....	126
31.3 Timeline of Artificial Intelligence.....	128

31.4 Artificial General Intelligence.....	131
32 Artificial Life.....	133
32.1 Artificial Life.....	134
33 AI and Creativity.....	136
33.1 Artificial Intelligence and Creativity.....	137
33.2 Curiosity in Artificial Intelligence.....	138
33.3 Artificial Intelligence in Video Games.....	140
34 Explainable AI.....	142
34.1 Explainable Artificial Intelligence.....	143
34.2 Artificial Intelligence Takeover.....	145
35 Ethical Aspects of AI.....	148
35.1 Ethical Aspects of Artificial Intelligence.....	149
35.2 Bio-Inspired Artificial Intelligence.....	152
35.3 Existential Risk from Artificial General Intelligence.....	155
35.4 Embodied Cognition.....	158
36 Next Generation Applications of AI.....	162
36.1 Next Generation Applications of Artificial Intelligence.....	163
36.2 Artificial Intelligence in Industry.....	164
36.3 Artificial Intelligence Industry in China.....	168
36.4 Artificial Intelligence in Hiring.....	169
37 Regulation of AI.....	171
37.1 Regulation of Artificial Intelligence.....	172
37.2 Philosophy of Artificial Intelligence.....	173
37.3 AI Ethics Fields' Approaches.....	175
37.4 EU AI Act - first regulation on AI.....	177
38 Additional Resources for Further Study.....	181
38.1 Additional resources for further Study I.....	182
38.2 Additional resources for further Study II.....	188
38.3 Additional resources for further Study III.....	196
39 Study Assignments Social and Ethical Aspects of AI.....	200
39.1 Serie 1.....	201
39.2 Serie 2.....	209
40 Notes and External Links.....	218
40.1 Resources.....	219
40.2 References.....	229

Introduction

Chapter **1**

1.1 Preface

1.1.1

This course “Social and Ethical Aspects of A.I.” has been conceived, just before the stormy introduction of ‘Open A.I.’ at the end of 2022. And a similar turbulent cascade of A.I. novelties may be expected in the coming decade as well. That implies that the mission of this course is just to make you aware of the main disputes going on so that its magnitude will no longer surprise you. Even we hope this course will harness you to relentlessly search for your own sources on how A.I. develops further, and at which point it touches your horizon of imagination; not only for intellectual purposes; also to continue undertaking your career plans and finding a good trade-off between commercial successes and recurring to what you find as contributing to the major values of life. These evangelical remarks can hardly be omitted as we see that A.I. touches the essence of what we tend to call ‘the existential basis of societal commitment and ethical dimensions’. In summary we may say that A.I. can no longer be seen as just instrumental to human survival. Just like human ‘free will’ and ‘art creations’ A.I. will stand on the shoulders of human intellect and may help humankind to find even more eternal values than we could even imagine from an anthropomorphic stage in awareness. In other words: A.I. might eventually protect earthly life from extinction and tumultuous geo-political conflicts as evolutionary ego-inertia tends to worry about local optima rather than global well-being.

From a more pragmatic point of view, we would like to encourage you to undertake your own scouting of news and literature on A.I. all the time: News rubrics like from the big ICT players, Google, Microsoft, Apple and Meta. They all resonate panic on what Open A.I. like in Chat-GPT brought along recently. So for you as students, it is a wake-up call that you cannot rely on just one or two main handbooks on A.I.; it is a day-to-day keenness on what appeared technically and socially. It urges you day-by-day to ask yourself: “What is in it for me? For my enterprise, my job, my family, my community? Etc. A first conclusion now (March 2023) is that Open A.I. will develop as background ‘intelligence’ to do your own jobs better and more underpinned from what is currently available around the globe.

Its immediate consequence is that many of the mentioned URLs in this course might have become obsolete or even disappeared even before the course unfolds before your eyes. This agility in information is an inevitable side effect from the fact that technologies and its societal commodities emerge so quickly. So, at this point in this course consolidation we recommend you to be an active and entrepreneurial learner: never stop exploring what appeared yesterday or even during the morning.

1.1.2

Finally: You will find at crucial locations in the course Graph representations of the main conceptual linkages that have been addressed in the paragraphs just before it. Its goal is to help you in your meta-cognitive awareness. These graphs have been generated through Graph-GPT (<https://graphgpt.vercel.app/>) and indeed demonstrate the use of Open A.I. in the generation of this course itself. The epistemic validity of these graphs are disputable, as also this element from Open A.I. Is still in its infant stage. However, we would like to provide you with all utilities that Open A.I. offers you at this very moment of course consolidation. Please enjoy the nature of this course, first of all by witnessing the next video on Science Fiction robot death machines.

Video:



"...A robotics researcher afraid of robots, Peter Haas, invites us into his world of understand where the threats of robots and artificial intelligence lie. Before we get to Sci-Fi robot death machines, there's something right in front of us we need to confront - ourselves. Peter is the Associate Director of the Brown University Humanity Centered Robotics Initiative. He was the Co-Founder and COO of XactSense, a UAV manufacturer working on LIDAR mapping and autonomous navigation. Prior to XactSense, Peter founded AIDG – a small hardware enterprise accelerator in emerging markets. Peter received both TED and Echoing Green fellowships. He has been a speaker at TED Global, The World Bank, Harvard University and other venues. He holds a Philosophy B.A. from Yale. This talk was given at a TEDx event using the TED conference format but independently organized by a local community. Learn more at <https://www.ted.com/tedx>"

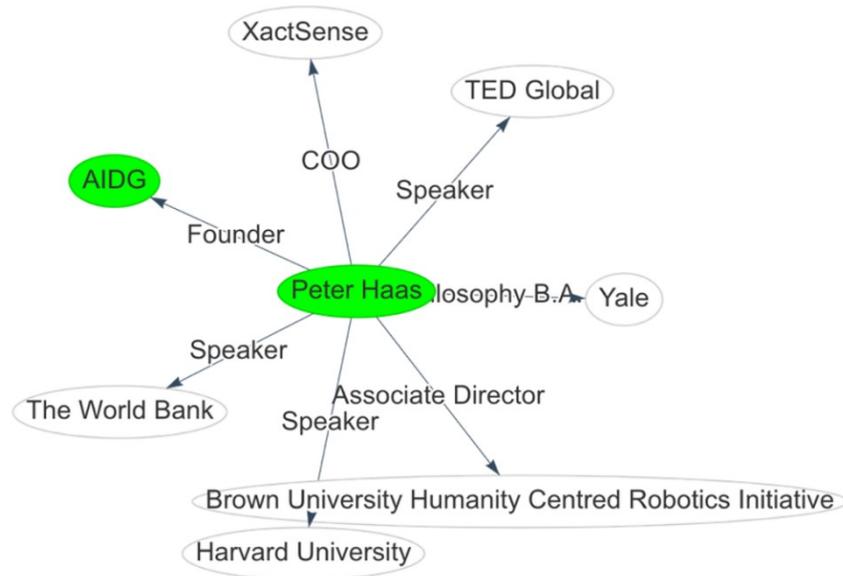


Figure: Peter Haas in his Various Roles in Robotics and Aligure

1.1.3

This video **AI** course module:



pleads for opening your mind and senses so that fundamental questions become welcomed rather than refuted. Also, this course module “Social and Ethical Aspects of Artificial Intelligence” has the goal to make you think of the human aspects of AI. As you will soon see in the first paragraphs of this course module, the social and **ethical aspects**



cannot be separated from the technical understandings that you will meet in the other modules of this course, like AI for Programming Languages, Data Processing, Machine- /Deep Learning, Language Processing, Cyber Security etc. A solid understanding both of the technology and AI’s human aspects will need an iterative approach: After each excursion, it will be needed to go back to the literature that you find in this module; But also, it is needed that you develop a life-long learning attitude so that during your later practice you will automatically tend to open websites with well-rubricized articles that have been carefully reviewed and validated. Especially in the topic of AI there is a tendency to intrude propaganda for certain power games; in short: ‘Fake news is a big threat to AI information as the tendency to influence public opinions is quite large’. For this sake of intellectual sanity, it was chosen to distil segments from Wikipedia. Each Wikipedia-driven paragraph has a link to the Wikipedia article, so that at later instances, the more recent version (based on the more recently validated understanding) will be reached. In order to let you read, select and jump dynamically, we have included the chosen Wikipedia paragraphs in this study texts, so that you can made a proper (efficient) decision to open the Wikipedia link, or just quickly scan the main lines of thought of the Wikipedia article via this course textbook. Another reason to include the Wikipedia texts themselves in this

course is that jumping to Wikipedia has the risk that you might digress too easily and lose the main line of this course module. Finally: At the end of this course module, twenty 'Study Assignments' have been formulated. Its goal is to bridge the prior literature with the more holistic questions that you may find in real-life situations like your Master- or Ph.D. study and not to forget, your career, where certainly AI will play a major role the coming decades.

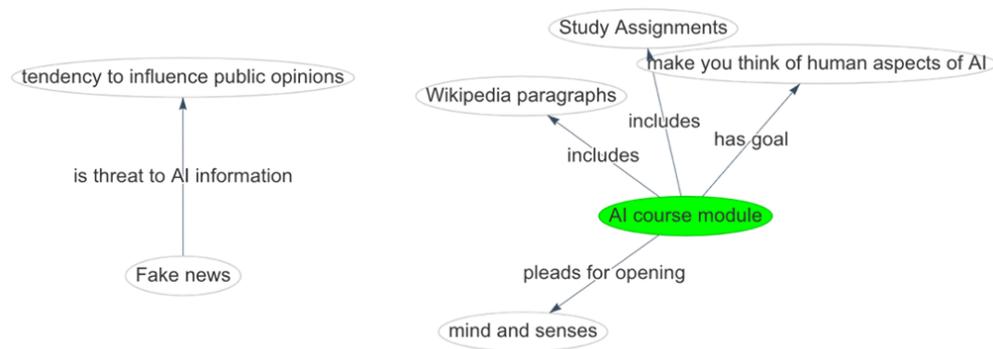


Figure: A.I. as a Threat? Follow and Exercise this Course's Assignment

1.1.4

Can you Remember? What technical aspects have been listed under AI traditionally? What are the major benefits for having this course module studied carefully?

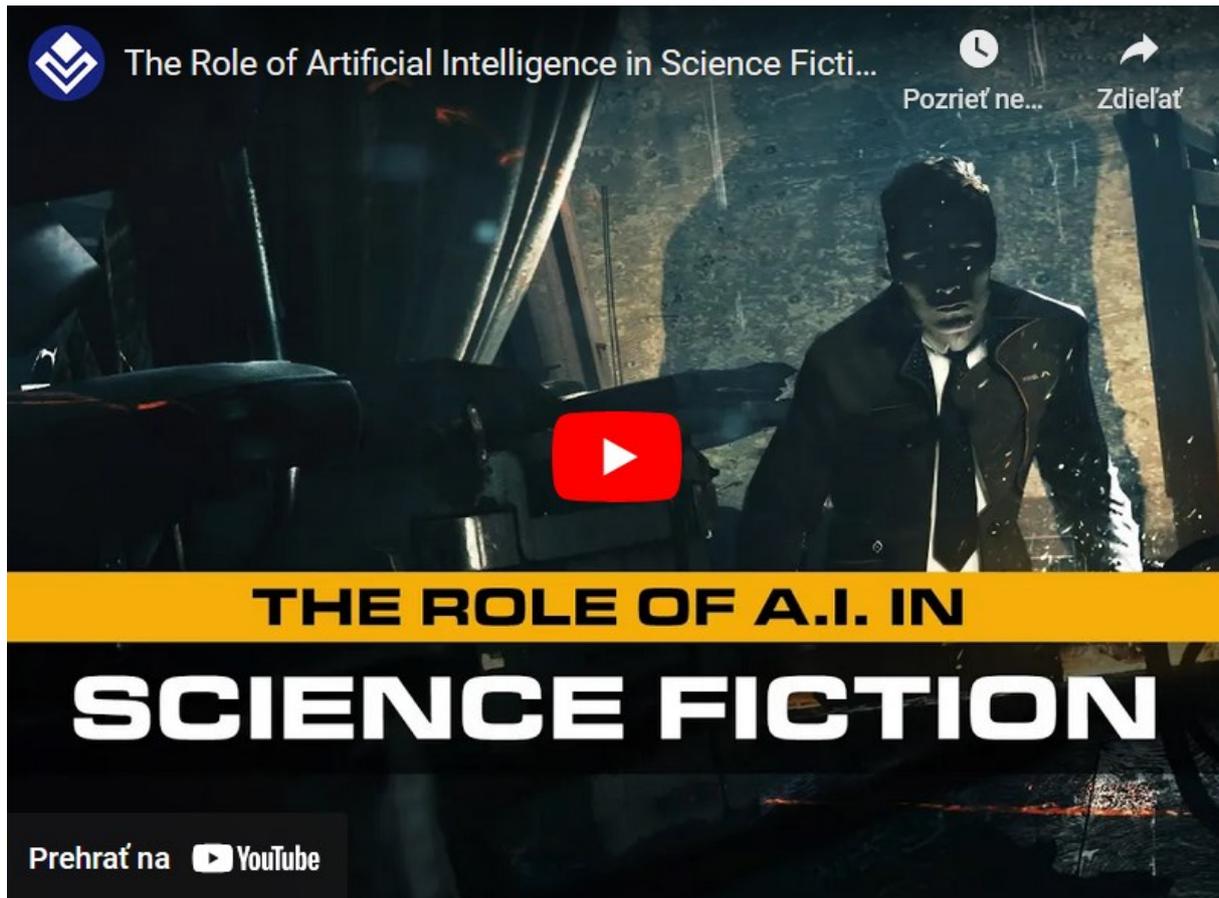
1.1.5

Question: Check the news media for recent events around A.I. Is it complementary to your earlier view on the essence of Artificial Intelligence? Feel welcome to share your findings with your peer students.

1.2 Introduction

1.2.1

The video **The Role of Artificial Intelligence in Science Fiction**



" 'Hello Future Me' dissects the structure and themes of artificial intelligence from *Detroit: Become Human*."

Please also check '[Mission AI in 12 Pictures](#)'. The recent report '[Mission AI: The New System Technology](#)' by Sheikh, Prins and Schrijvers; Springer Verlag, 2023, mentions five superordinate tasks when it comes to embedding AI technologies in societal contexts:

1. Demystification: Tackling overly optimistic and pessimistic images and learning to focus on the right questions.
2. Contextualization: Making the technology work in practice by creating an enabling socio-technical ecosystem.
3. Engagement: Democratizing the technology by involving relevant actors, in particular civil society.
4. Regulation: Developing appropriate regulatory frameworks that safeguard fundamental rights and values in the long-term.
5. Positioning: Investing in competitiveness and assuring security in an international context.

 1.2.2

- Make learning about AI and its application an explicit goal of governmental policy.
- Stimulate the development of 'AI wisdom' amongst the general public, beginning by setting up algorithm registers to facilitate public scrutiny.
- Explicitly choose an 'AI identity' and investigate in which domains changes in the technical environment are required to realize this.
- Enhance the skills and critical abilities of individuals working with AI and establish educational training and forms of certification to qualify people.
- Strengthen the capacity of organizations in civil society to expand their work to the digital domain, in particular with regard to AI.
- Ensure strong feedback loops between the developers of AI, its users, and the people that experience its consequences.
- Connect the regulatory agenda on AI to debates on the principles and organization of the 'digital environment' and develop a broad strategic regulatory agenda.
- Use regulation to actively steer developments of surveillance and data collection, the concentration of power, and the widening gap between the public and private sector in the digital domain.
- Bolster national competitiveness through a form of 'AI diplomacy' that is focused on international cooperation, specifically within the European Union.
- Know how to defend yourself in the AI era; strengthen national capacities to combat both information warfare and the export of digital authoritarianism.
- Finally, the report formulates a recommendation to address the institutional aspects of embedding AI within society: Establish a policymaking infrastructure for AI, starting with an AI coordination centre that is embedded into the political process. For the sake of this course module, it means that you as student and young career candidate, you need to become keen on following every-day news on AI; what game-changers are ongoing, and what does it mean for your career.



Figure A.I. in its Current Societal Contexts

Historical Context

Chapter **2**

2.1 Historical context

2.1.1

'Artificial Intelligence' started in the late fifties as topic for 'Science Fiction' and kept as academic outskirts during the next forty years until the 'Fifth Generation Project' caught momentum under the prestige of 'Expert Systems' in the nineteen eighties and ninetens.

1. The idea was: "If we convert human knowledge and thinking into machine-readable code, we can make machines even smarter than humans". Not as a surprise, it was in the domains of medical reasoning, seismic pattern analysis for finding geo-resources, and ... for making 'intelligent instructional systems that could trace students' misconceptions and optimize learning sequences.
2. More recently, two significant milestones were met: 'Deep Blue' as attempt to let computer chess playing compete the best human player Kasparov at that moment. Deep Blue won, however Kasparov complained that it was not a fair game, as Deep Blue had access to a large external library of winning moves in the past. So, in fact, Kasparov pleaded to play against Deep Blue with the help of a chess computer for himself. Indeed, it was the proof that from now on, the machine was the winner in chess playing. Even more remarkable was 'Alpha Go' about ten years after 'Deep Blue'; (Dong, X., Wu, J. and Zhou, L. 2017.) Alpha Go was a computer program, not primarily for winning the Go game; it started searching for heuristics ('rules of thumb') in the immense combinatorics of the Go middle game. So far, no human had succeeded to distil the hidden rules that can help to reduce the millions of combinations after the first steps were made by both players. This was an enormous step forward: the computer helping the human player to become a better player than ever before.
3. Nowadays, AI has appeared everywhere in life: from personal assistants for smartphones to chatbots for customer support, from recommending entertainment to predicting crime, and from facial recognition to medical diagnoses.

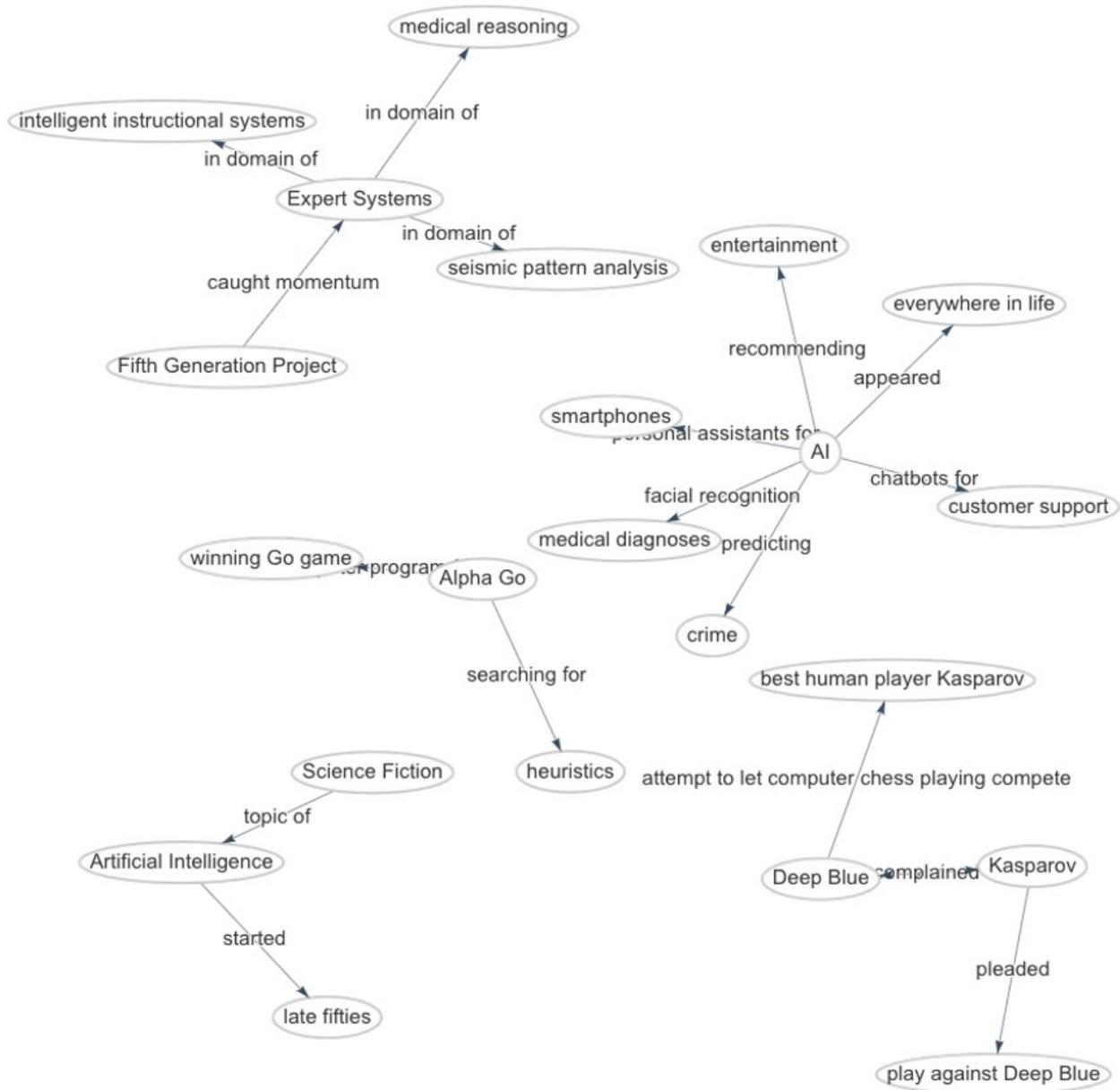


Figure: Trickling down from Medical to Consumer Goods into a new Science Fiction

 2.1.2

Can you Remember? What were the first three domains that were encoded into early AI systems? Which two games were equipped with AI and what fundamental understanding did they bring?

 2.1.3

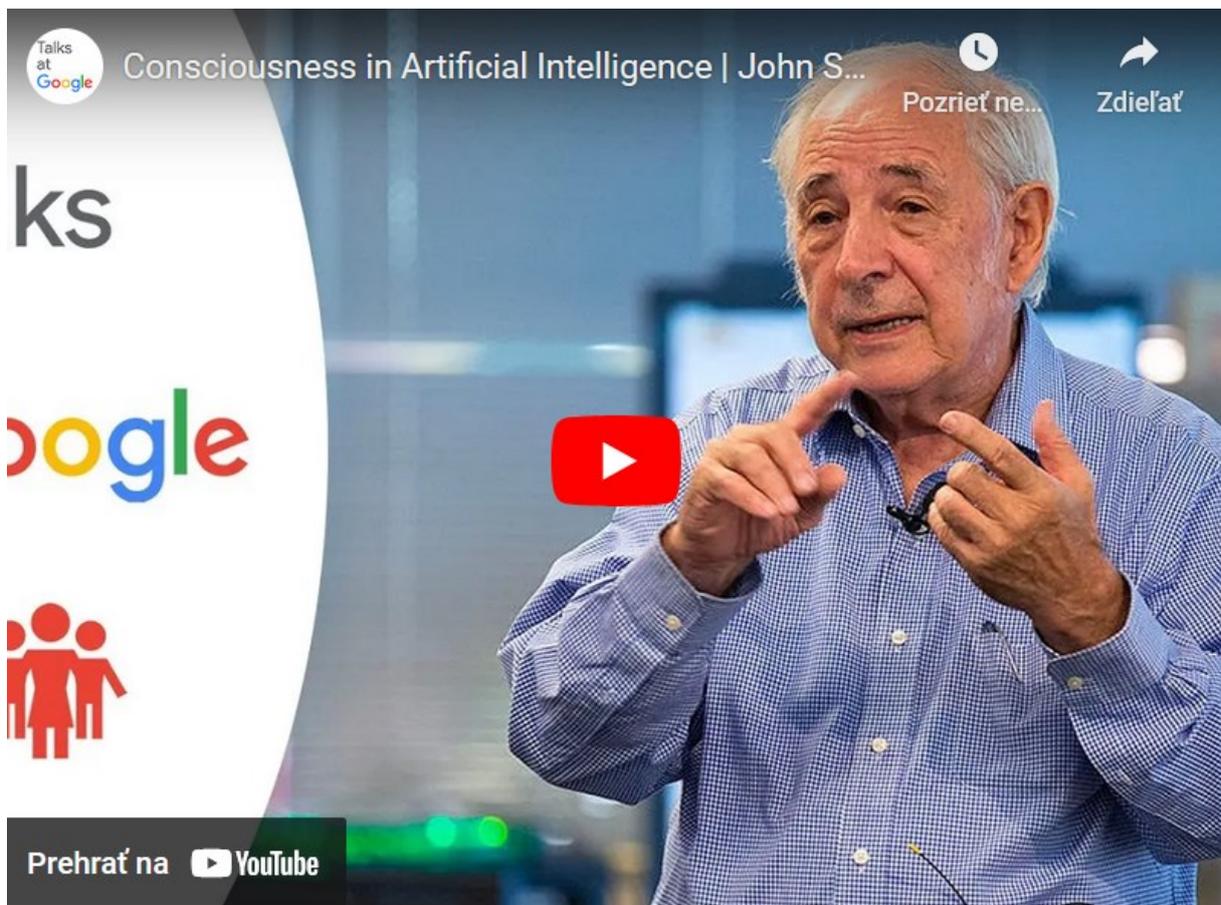
Question: In this course module there is ample awareness on how citizens may become victim to AI mechanisms and will soon loose part of their privacy and autonomy. Have you thought about your own data in the broader context of AI like social media, cameras on the streets and squares? If so, please look back to the situation of about one year ago and compare the difference.

2.2 Trust and Consciousness in AI

2.2.1

Wikipedia: [Trust and Consciousness in AI](#)

Video: **Consciousness in Artificial Intelligence**



“ John Searle is the Slusser Professor of Philosophy at the University of California, Berkeley. His Talk at Google is focused on the philosophy of mind and the potential for consciousness in artificial intelligence. This Talk was hosted for Google's Singularity Network. John is widely noted for his contributions to the philosophy of language, philosophy of mind and social philosophy. Searle has received the Jean Nicod Prize, the National Humanities Medal, and the Mind & Brain Prize for his work. Among his notable concepts is the "Chinese room" argument against "strong" artificial intelligence.”

It has been known for decades that people are not very well equipped for monitoring automated processes. That problem is simple and fundamental: the more automated the process is and runs smoothly, the more difficult it becomes for people to remain actively involved in the process. If something then goes wrong and adjustments have to be made, then at that moment the human operator is not well informed about the current situation and is also no longer practiced in manually adjusting the process.

The effect that people are too accommodating, too trusting is referred to as complacency. This compliance is increased if someone does not see the space to search further. Especially in busy times and under stress, the operator is more likely to accept the suggestions of the system for reliable. While in quieter times the same operator would check more often whether the system responds correctly to data from the process. In collaboration with automated systems, people are easily side-lined. These problems are usually referred to by the English term: 'out-of-the-loop'. Or also with the solution-oriented formulation: 'Human in the loop'. It is necessary to keep humans involved in the process, at least if you want man to remain vigilant with the aim of intervening and adjusting if necessary. If people are side-lined, they lose sight of the current situation and developments in the process. Knowing and staying informed about the current situation and developments is referred to by the term 'situational awareness' or more often by the English term 'situational consciousness'.

Situational consciousness is essentially the current mental model that a human being has of the situation and the developments in it. This model is essential for making decisions and adjusting a process. We distinguish 3 levels of situational consciousness

1. Knowing the current values of the critical parameters in the process.
2. Understand the meaning of these values and especially the combined influence of these values on the task objective.
3. Be able to estimate how the process will develop in the near future. What if automation puts people out of the game, but human surveillance is still necessary and manual intervention is also necessary? Many of the problems outlined earlier arise from the starting point: 'Automating what can be automated.'

2.2.2

Can you Remember? What wanted John Searle clarify with his 'Chinese Room' metaphor? What is the meaning of 'keeping human operators in the loop?' Can you repeat the three main elements of 'situational awareness'?

2.2.3

Question: Complacency is often mentioned as basic attitude that makes citizens vulnerable: Policy makers, politicians and employers are then tempted to take a paternalistic stand and defend their tutelage by saying that citizens 'need guidance' in order not to become victim of naivety and backwardness. What is your opinion about how to make the right balance between tutelage and staying in a complacent ignorance?

2.3 Rapid Course Rational

2.3.1

This course module intends to stimulate your active learning and updating in the field of 'AI and its social and ethical aspects'. This domain contrast with the pure technological perspectives like 'specifying systems' behaviour', 'coding', 'testing and debugging' before a real system is supposed to work. In contrast to these technological demands and constraints, the social and ethical concerns need:

1. Critical Thinking.
2. Out of the box Thinking in order to find Creative Solutions and Unorthodox Problem Solving.
3. The Willingness and Skills to Listen and Read other Scholars' Argumentations in an Empathic way.
4. To develop a constant Eagerness to Screen News Channels and Societal Dialogues on AI via the Media Channels.

In order to prepare these needed skills and attitude, in this course there will be blogging facilities in order to sharpen the dialogues between students. The goal will be to not land in polarization and caricatures. Even, it will be encouraged that you raise your own questions and discussions. Taking this in mind, both the two types of questions in this course module, have a catalytical goal. It is not meant to discriminate 'correct' from 'faulty' answers. The quality for finally assessing your grade will be based upon

1. The degree of your alertness in bringing forward authentic ideas and the degree of empathy during your interactions with your peer students.
2. Finally the degree that you manage to integrate notions from social and ethical nature in the other modules of this course.

If you have doubts on how to handle these criteria, please consult the teacher or the mentor of this course.

2.3.2

TED video: **Why you think you're right -- even if you're wrong**



".... Perspective is everything, especially when it comes to examining your beliefs. Are you a soldier, prone to defending your viewpoint at all costs – or a scout, spurred by curiosity? Julia Galef examines the motivations behind these two mindsets and how they shape the way we interpret information, interweaved with a compelling history lesson from 19th-century France. When your steadfast opinions are tested, Galef asks: "What do you most yearn for? Do you want to defend your own beliefs, or do you want to see the world as clearly as you possibly can?""

The goal of this course module is to make you aware of the rapid technological developments that inevitably bring multiple risks and challenges, leading to policy debates and understand the needed regulatory frameworks. The public worries nowadays are that AI may overpower human integrity, privacy and even overtake political decisions on peace and war. And also, that AI may become a tool that make the rich richer and the poor poorer. The flip side of the worries is that AI may make education, care and industries more effective and efficient. It can be seen in recent opinions, that the cross-fertilization of AI-empowered sectors like the climax up to Society 5.0 in combination with Industry 4.0 as in Figure below.

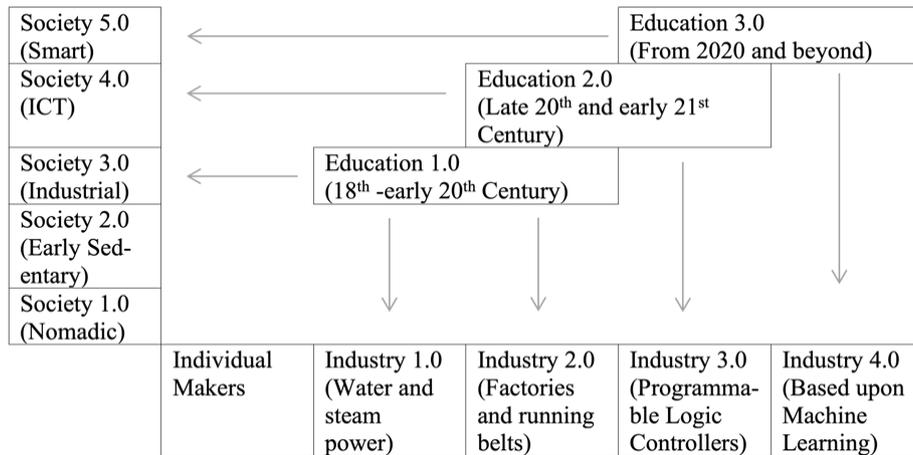


Figure: Society 5.0 and Industry 4.0 as Drivers for Education 3.0

 2.3.3

Can you Remember? What is essential for the transition Society 4.0 to 5.0? A for the transition Industry 3.0 to 4.0?

 2.3.4

Question: If you look to Figure 1 above; What role do you estimate Education plays in the cascade from Society 4.0 to Society 5.0 and Industry 3.0 to Industry 4.0? If you have no clue, feel happy to consult the web and please watch the next video.

Industry 4.0

Chapter **3**

3.1 Industry 4.0

3.1.1

Wikipedia: [Industry 4.0](#)

Video: **What Is Industry 4.0?**



In this video Bernard Marr explains, in very simple terms what Industry 4.0 is, also referred to as the 4th Industrial Revolution. If you would like more information on this topic, please feel free to visit his website! He writes articles every week on various topics such as Big Data, AI and Key Performance Indicators.

A buzz word that stands for the current automation wave in the industry, characterized by AI and IoT. It is better to say that there is an ongoing trend than that there is a fourth industrial revolution, as suggested by the term Industry 4.0.

Three previous industrial revolutions were:

1. artificial force (steam engines),
2. easy distribution of energy in the form of electricity and production in series work instead of customization per example, and
3. artificial decision-making power, digitization, automation.

A robot is a programmable machine that can perform physical actions. On the shop floor, these machines often have the form of a movable arm with a gripper or tools such as 'hand' and various degrees of freedom to move. But also numerically controlled machining machines (lathe, saw, milling cutter, laser cutter) are robots. Robots work in a shielded environment for safety reasons. A cobot is a collaborative robot, a robot with such protections that they are allowed to work (together) in the environment of people. The following terms are more or less in order of control level. The highest level is management and planning.

The lowest level are the sensors and actuators that are in direct connection with the physical production process.

For more details see the journal: "[Sensors and Actuators A: Physical brings together multidisciplinary interests in one journal entirely devoted to disseminating information on all aspects of research and development of solid-state devices for transducing physical signals.](#)"

3.1.2

Can you Remember? What is a 'cobot'? What 'levels of control' can we distinguish in roboting?

3.1.3

Question: Each upward transition in Industry brings the need to upgrade workers' qualifications. We have seen that craftsmanship in the transition Industry 3.0 to 4.0 has urged workers to incorporate 'computational' and 'algorithmic' thinking instead of having a sharp eye for size and material responses. What typical new demands for industrial workers do you estimate to arrive after the transition from Industry 4.0 to 5.0?

Citizens 2023

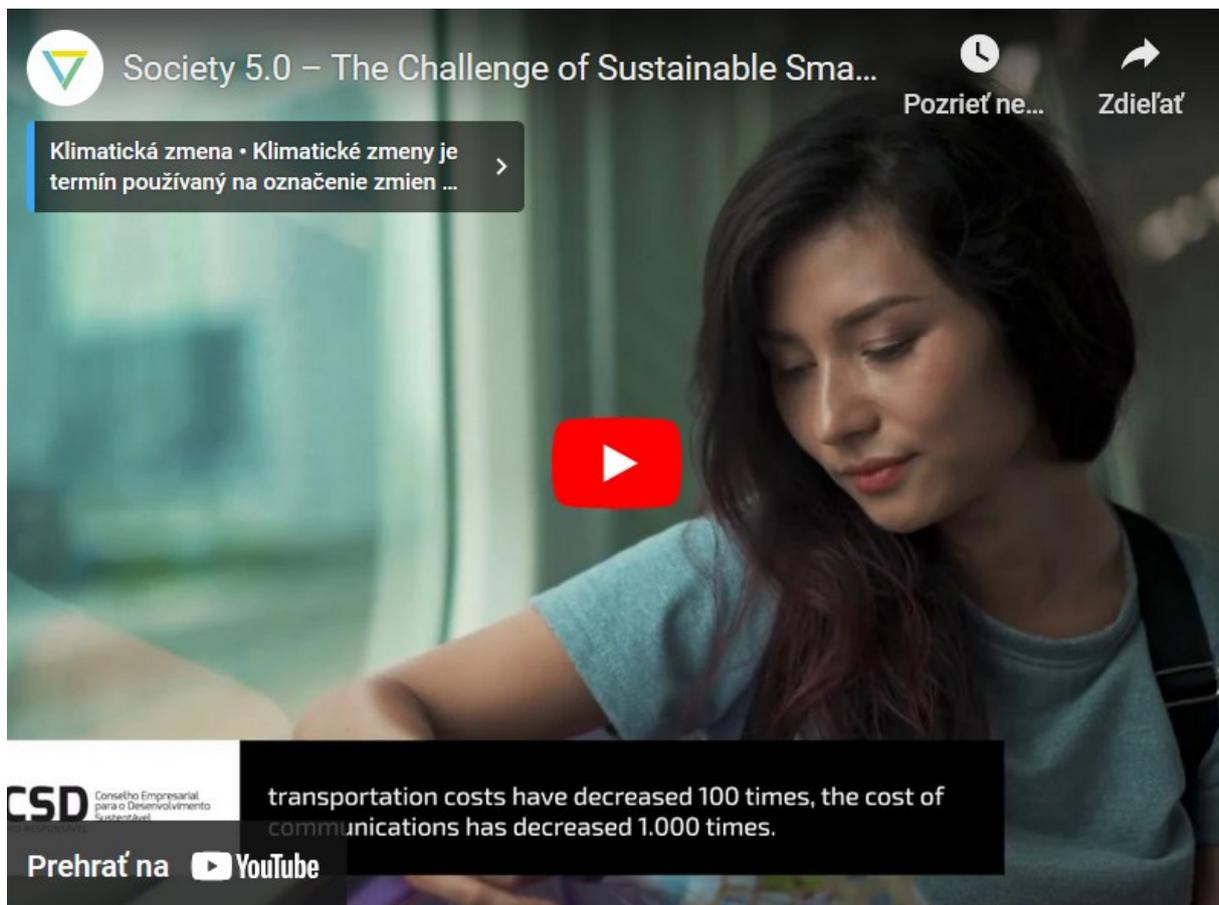
Chapter **4**

4.1 Citizens 2023

4.1.1

Wikipedia: [Citizens 2023](#)

Video: **Society 5.0 – The Challenge of Sustainable Smart Societies**



The same may occur for instance for weapon industry in the intersection between social intelligence and drone technologies. Social intelligence has been lifted beyond our initial imagination: At the start of social media in 2010, no idea emerged that anno 2023 citizens would feel threatened by the amount of privacy that would already be incorporated in systems like Facebook, Twitter, TikTok and by Google. Still, we are enthusiastic what AI-supported systems can offer to medical diagnoses and even surgical interventions.

Also, AI has entered the world of education. 'Intelligent', 'adaptive' and 'personalized' learning systems are increasingly being developed by the private sector for deployment in schools and universities around the world, creating a market expected to be worth US\$6 billion in 2024 (Bhutani and Wadhvani, 2018). Indeed, also the application of AI in educational contexts raises profound questions – for example about what should be taught and how, the evolving role of teachers, and AI's social and ethical implications. It will become clear to you in this course module, that if we

give full autonomy to technology, we may soon see increased social/economic disbalance, social revolt against the power of larger commercial entities and the call for state regulations based upon democratic outcomes. (Smith and Neupane, 2018, p. 12)

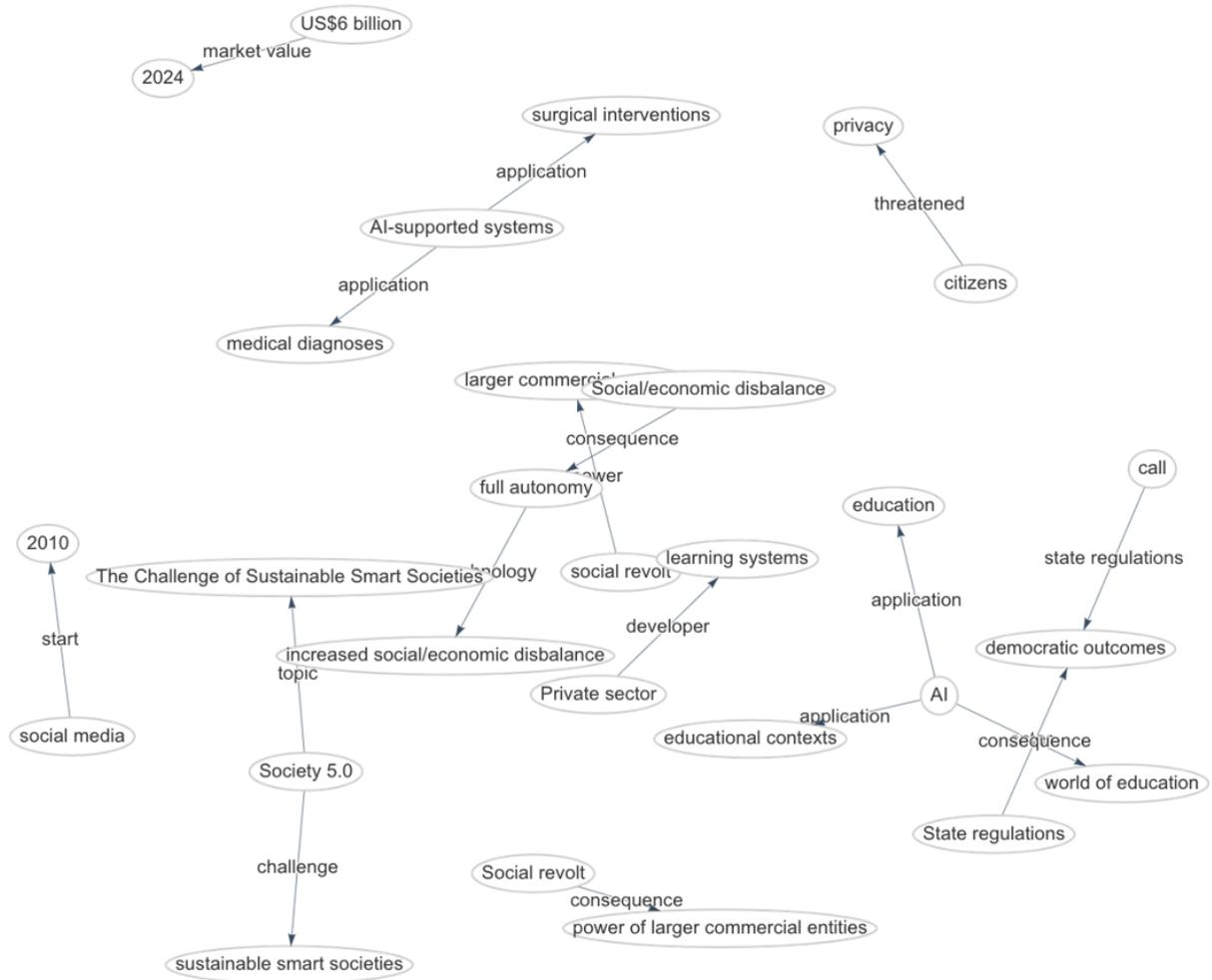


Figure: What are the 'Need' Spheres for A.I.?

 4.1.2

Can you Remember? What side effects can we expect if full priority to technological autonomy is allowed?

 4.1.3

Question: Given that Society 5.0 brings citizens in the condition of 'Smart Cities' with abundant sensors and AI for optimizing traffic streaming and reasoning about safety, environmental health etc. What transition do you estimate education to go through in terms of curricular content and instructional methods?

AI Needs a Blend of Disciplines and Values

Chapter **5**

5.1 AI needs a Blend of Disciplines and Values

📖 5.1.1

Wikipedia: [AI needs a Blend of Disciplines and Values](#)

Our best practice business frameworks, templates, and tools are of the same calibre as those produced by top-tier management consulting firms, like McKinsey, BCG, Bain, and Accenture. Most were developed by seasoned executives and consultants with 20+ years of experience. The Flevy Marketplace represents the largest library of best practice business documents available online--with over 5,000+ documents (including frameworks & methodologies, presentation templates, financial models, and more), spanning over 350+ management topics (<https://flevy.com/business-toolkits>).”

The three major ingredients for AI are: culture, nature and technology

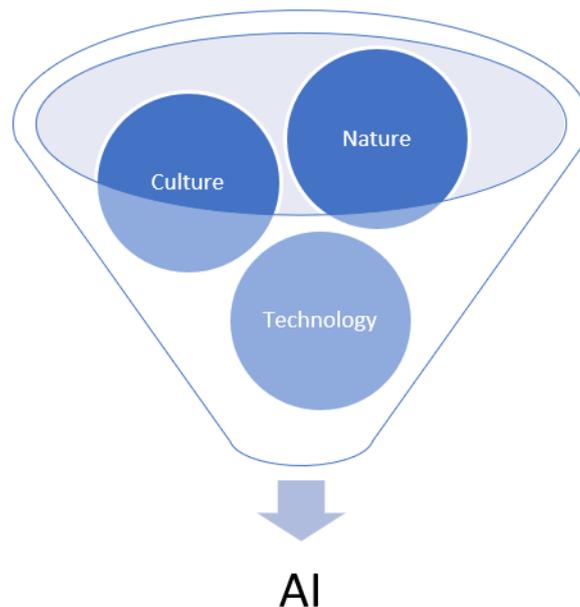


Figure: How A.I. derives from Culture, Nature and Technology?

📝 5.1.2

The entrance of technology is not a recent guest: It stems upon the arrival of using an object as a tool, controlling fire, taming animals, and not at least: using ICT for channelling learning and thinking via simulations and programmed instruction. The fact that media instigated. The orientation towards what AI will bring us, and how we

can prevent from its misuse soon gets interwoven with philosophical questions on ‘what intelligence is’ and ‘can computers also become wise and responsible?’ etc.

As we saw in the example of Alpha Go, it is maybe not the best to oppose human- to machine intelligence, but rather say that the combination of the two is what we may expect a future break-through. ‘Machine learning’ has become the major epitome of what we call ‘AI’. It rests upon ‘deep learning’ and ‘neural networks’. Almost all the world’s big technology companies, and many others, now offer sophisticated ‘AI-as-a-service’ platforms, some of which are open-source.

These provide various AI building-blocks that the developers can implement without having to write AI algorithms from scratch. Table 1 lists the more dominant attempts of technology providers and the AI service platforms for 3rd party memberships.

Table: Companies offering AI Services

Technology Company	AI as a Service Platform	Company’s Description
1. Alibaba	Cloud	Cloud-based AI tools to support the demands of businesses, websites, or applications: https://www.alibabacloud.com
2. Amazon	AWS	Pre-trained AI service for computer vision, language, recommendations, and forecasting. It can quickly build, train and deploy machine learning models at scale or build custom models with support for all the popular open-source frameworks: https://aws.amazon.com/machine-learning
3. Baidu	EasyDL	Supports customers to build high-quality customized AI models without having to code: https://ai.baidu.com/easydl
4. Google	TensorFlow	An end-to-end open-source platform for machine learning, including an ecosystem of tools, libraries and community resources that enables researchers to share the state-of-the-art in machine learning and developers to easily build and deploy machine-learning-powered applications: https://www.tensorflow.org
5. IBM	Watson	Allows users to bring AI tools and apps to the data wherever it resides regardless of the host platform: https://www.ibm.com/watson
6. Microsoft	Azure	Includes more than 100 services to build, deploy and manage applications: https://azure.microsoft.com
7. Tencent	WeStart	Maps AI capabilities, professional talent and industry resources to support the launch or enhancements of start-ups. It connects industry partners, disseminates, and applies AI technology in different industry sectors: https://westart.tencent.com/ai

 5.1.3

In summary we may conclude that already machines become capable to imitate aspects of human intelligence like learning, problem solving, pattern recognition, reasoning, etc. Each of these human cognitive ‘faculties’ have only partially been described and explained in psychology.

If we talk about social intelligence, emotion, intuition and creativity, there is even a wider gap to be taken; AI will enforce us to become more explicit on what we mean by them. AI finally will help humans to understand nature better and better; Similar to what was already offered by science, cognitive- and behavioural theories the past two centuries. Based upon mathematics, predicate logic and emulating neural networks, we already progressed a lot.

Real-world applications of AI are becoming popular and its pro- and contra ponents mention the risks of full-automatic translation of natural language in sensitive cases like jurisprudential and political discourses.

The same is with face-recognition and medical prescriptions and diagnosing foetal aberrations. It was found out, that when medical experts combine AI-based imaging technologies together with radiologists, the combination of the AI technology and the radiologist outperforms either the AI or the radiologist by themselves.

The growth of AI and automated processes often creates concerns that the human touch will be removed from the health-care delivery process. What the industry is finding, however, is the opposite is becoming true: AI can extend the resources and capabilities of overworked health-care professionals and vastly improve processes for medical interventions.

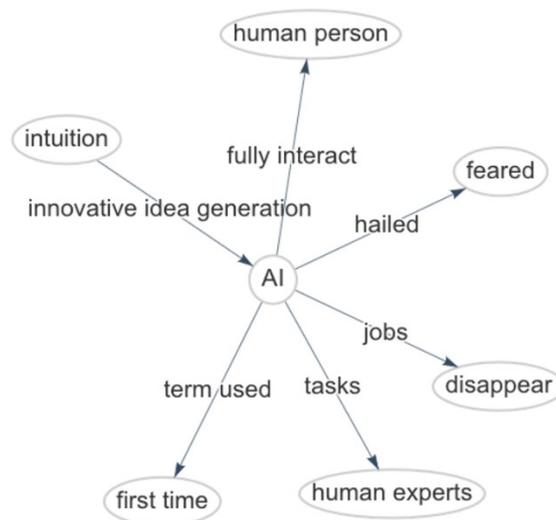


Figure: Factors in the Adoption of A.I.

 5.1.4

Remember: Which AI service platforms can you mention? Which large AI-dependent enterprise is missing remarkably?

 5.1.5

Question: Try out each of the seven given AI platforms; What are its first glance strengths and weaknesses? What human task would you favour to support with AI? How would you divide its human- versus the machine component?

Current State of the Art in OpenAI

Chapter **6**

6.1 Current State of the Art in OpenAI

📖 6.1.1

[OpenAI link](#)

According to [Wikipedia](#):

" .. OpenAI is an [artificial intelligence](#) (AI) research laboratory consisting of the for-profit corporation OpenAI LP and its [parent company](#), the non-profit OpenAI Inc. The company, considered a competitor to [DeepMind](#), conducts research in the field of AI with the stated goal of promoting and developing [friendly AI](#) in a way that benefits humanity as a whole. The organization was founded in [San Francisco](#) in late 2015 by [Elon Musk](#), [Sam Altman](#), and others, who collectively pledged US\$1 billion. Musk resigned from the board in February 2018 but remained a donor. In 2019, OpenAI LP received a US\$1 billion investment from [Microsoft](#)..."

📖 6.1.2

On 2022 December 10 the public announcement of [OpenAI](#) became in the news. In order to explore Open AI's functionality, please subscribe and authorize yourself via <https://beta.openai.com/>. The OpenAI API can be applied to virtually any task that involves understanding or generating natural language or code. We offer a spectrum of [models](#) with different levels of power suitable for different tasks, as well as the ability to [fine-tune](#) your own custom models. These models can be used for everything from content generation to semantic search and classification. At the moment of conceiving this course module, Open AI's its components are:

1. [Image Generation](#): The Images API provides three methods for interacting with images: Creating images from scratch based on a text prompt; Creating edits of an existing image based on a new text prompt; Creating variations of an existing image; This guide covers the basics of using these three API endpoints with useful code samples. To see them in action, check out our [DALL·E preview app](#).
2. [Text Completion](#): The [completions](#) endpoint can be used for a wide variety of tasks. It provides a simple but powerful interface to any of our [models](#). You input some text as a prompt, and the model will generate a text completion that attempts to match whatever context or pattern you gave it. For example, if you give the API the prompt, "As Descartes said, I think, therefore", it will return the completion " I am" with high probability.
3. The best way to start exploring completions is through our Playground. It's simply a text box where you can submit a prompt to generate a completion. To try it yourself, [open this example in Playground](#):
4. [Code Completion](#): The [Codex model series](#) is a descendant of our [GPT-3 series](#) that's been trained on both natural language and billions of lines of

code. It's most capable in Python and proficient in over a dozen languages including JavaScript, Go, Perl, PHP, Ruby, Swift, TypeScript, SQL, and even Shell. During this initial limited beta period, Codex usage is free. [Learn more](#).

5. [Embeddings](#): An embedding is a special format of data representation that can be easily utilized by machine learning models and algorithms. The embedding is an information dense representation of the semantic meaning of a piece of text. Each embedding is a vector of floating point numbers, such that the distance between two embeddings in the vector space is correlated with semantic similarity between two inputs in the original format. For example, if two texts are similar, then their vector representations should also be similar.
6. [Job Finding](#): Developing advanced artificial intelligence systems requires people from a wide range of disciplines and backgrounds. At OpenAI, we're always looking for curious minds that want to help us create safe and beneficial AI.
7. And many more to come soon.

6.1.3

According to [OpenAI](#): “ .. DALL·E is a 12-billion parameter version of [GPT-3](#) trained to generate images from text descriptions, using a dataset of text–image pairs. We’ve found that it has a diverse set of capabilities, including creating anthropomorphized versions of animals and objects, combining unrelated concepts in plausible ways, rendering text, and applying transformations to existing images.”



Figure: A.I. as Engine for Pictorial Analogies

6.1.4

“ .. GPT-3 showed that language can be used to instruct a large neural network to perform a variety of text generation tasks. [Image GPT](#) showed that the same type of neural network can also be used to generate images with high fidelity. We extend these findings to show that manipulating visual concepts through language is now within reach. [ChatGPT](#) offers the Optimization of Language Models for Dialogue: “ .. We’ve trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer follow-up questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests. ChatGPT is a sibling model to [InstructGPT](#), which is trained to follow an instruction in a prompt and provide a detailed response.” As example, when we query “ideas for a 10-year old's birthday:” the set of suggestions is:

1. “ .. Host a scavenger hunt with clues and riddles that lead to different locations around the neighborhood or house. The final clue could lead to a special surprise or gift.
2. Set up a DIY spa day with facemasks, nail polish, and other relaxing activities. You could even add in some healthy snacks and mocktails for a fun and pampering experience.
3. Plan a movie night with the birthday child's favorite films, popcorn, and other snacks. You could also make a red-carpet entrance and have everyone dress up in their favourite movie-themed costumes.
4. If the weather permits, have a backyard camping adventure. Set up tents, roast marshmallows over a fire, and tell spooky stories.
5. Plan a trip to a local amusement park or trampoline park for a day of fun and excitement. This can be a great way for the birthday child and their friends to burn off some energy and have a blast.

Overall, the key is to plan activities that the birthday child will enjoy and that will create lasting memories... “

6.1.5

In summary, Open AI mentions on its [website](#):

“ .. OpenAI is an AI research and deployment company. Our mission is to ensure that artificial general intelligence benefits all of humanity. OpenAI’s mission is to ensure that artificial general intelligence (AGI)—by which we mean highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity. We will attempt to directly build safe and beneficial AGI, but will also consider our mission fulfilled if our work aids others to achieve this outcome..”

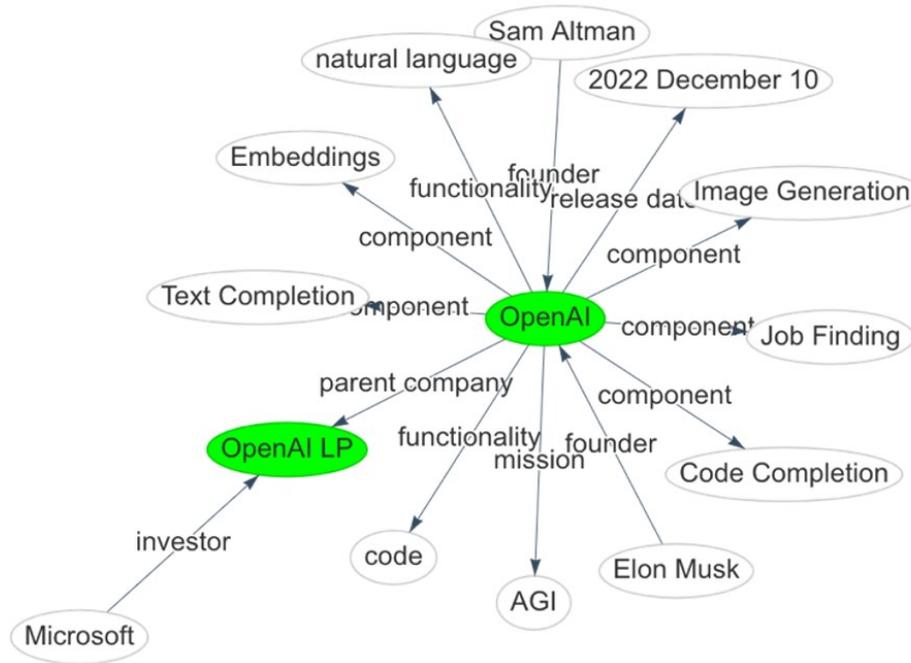


Figure: How Open A.I. created an Avalanche among the Big ICT Players

6.2 ChatGPT: Optimizing Language Models for Dialogue

📖 6.2.1

[About ChatGPT](#)

"We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer follow-up questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests. ChatGPT is a sibling model to InstructGPT, which is trained to follow an instruction in a prompt and provide a detailed response."

Try it now at www.chat.openai.com.

📝 6.2.2

Question: After having tried out Open AI and having read its [Charter \(https://openai.com/charter/\)](https://openai.com/charter/); What is the difference between consulting Google, Quora or Watson and OpenAI in its approach and achievement.

Adjacent Fields for Applying AI

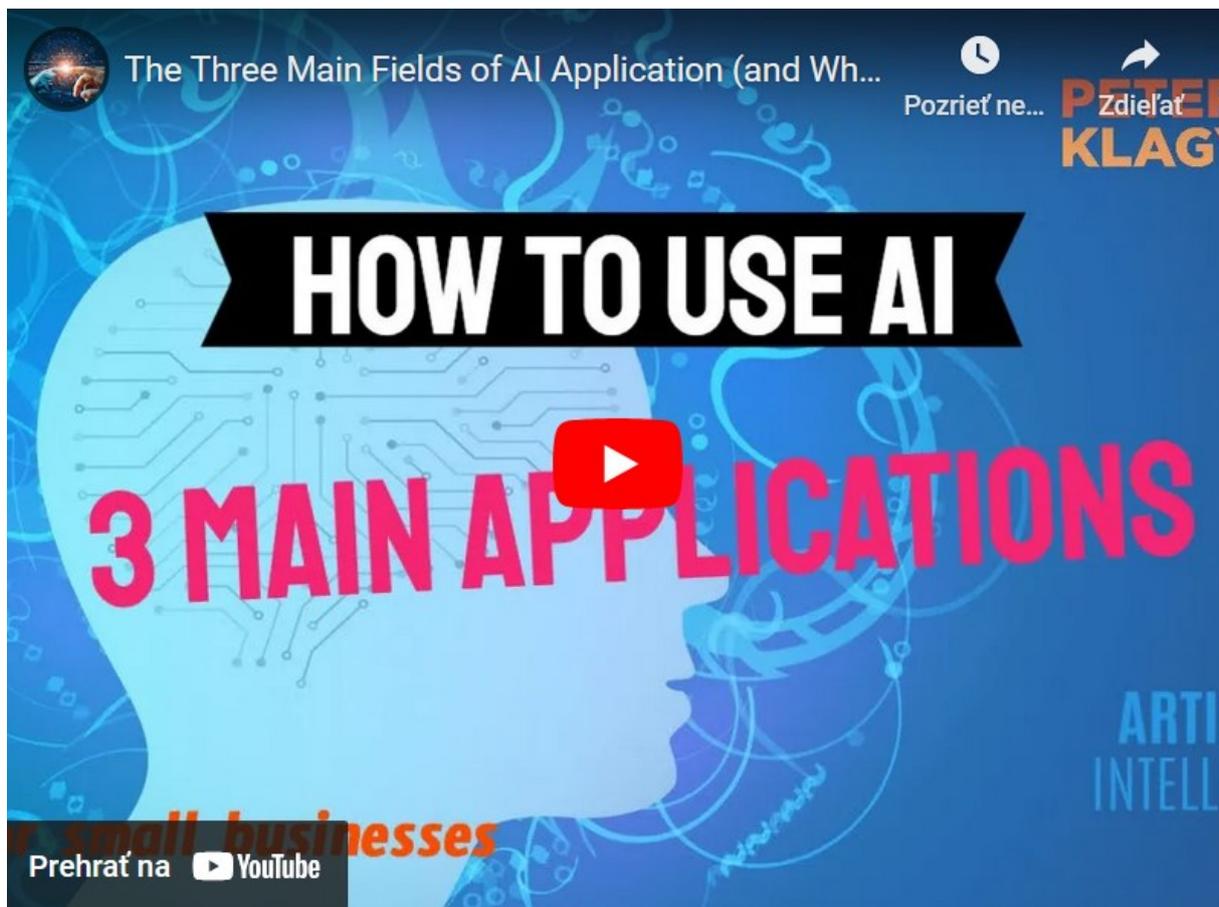
Chapter **7**

7.1 Adjacent Fields for Applying AI

7.1.1

Wikipedia: [Adjacent Fields for Applying AI](#)

Video: **The Three Main Fields of AI Application (and Why You Need To Be There**



" The three main fields of Artificial Intelligence Application I really think all businesses must use the power of AI. Artificial Intelligence is not the future anymore. It is here, it is happening NOW! There are three main fields where every business can benefit from AI. Automation/optimization - reduce cost - increase revenue. The tool mentioned in the video: Conversion.ai - Use AI to write proven, high converting copy for increased conversions and higher ROI. Instantly generate high-quality copy for ads, emails, websites, listings, blogs, and more..."

- Auto-journalism AI agents continually monitoring global news outlets and extracting key information for journalists, and also automatically writing some simple stories.
- AI for legal services like providing automatic discovery tools, researching case law and statutes, and performing legal due diligence.
- AI weather forecasting Mining and automatically analyzing vast amounts of historical meteorological data, in order to make predictions;

- AI fraud detection Automatically monitoring credit card usage, to identify patterns and anomalies (i.e., potentially fraudulent transactions);
- AI-driven business processes like for example, autonomous manufacturing, market analysis, stock trading, and portfolio management;
- Smart cities using AI and the interconnected Internet of Things (IoT) to improve efficiency and sustainability for people living and working in urban settings; and AI robots' Physical machines that use AI techniques, such as machine vision and reinforcement learning, to help them interact with the world.

While each of these examples have significant positive potential for society, we should not neglect to point out that other applications of AI are more controversial. Two examples are: Autonomous Warfare Weapons, drones and other military equipment that function without human intervention; and deep-fakes Automatic generation of fake news, and the replacement of faces in videos so that politicians and celebrities appear to say or do things they never said or did. In addition, we should also be careful when evaluating many of the dramatic claims made by some AI companies and the media. To begin with, despite headlines announcing that AI tools are now 'better' than humans at tasks such as reading texts and identifying objects in images, the reality is that these successes are only true in limited circumstances – for example, when the text is short and contains enough required information for inference to be unnecessary. Current AI technologies can also be very brittle. If the data is subtly altered, for example, if some random noise is superimposed on an image, the AI tool can fail badly (Marcus and Davis, 2019). For a more in-depth diachronic orientation of how AI ambitions shifted in time, please consult '[AI at a Crossroads](#)' by Luc Steels.

7.1.2

Can you Remember? Which three major benefits from AI have been listed in the video by Peter Klagyivik? Which two controversial societal aspects of AI have been mentioned?

7.1.3

Question: Thinking about AI-driven weapons and robots like drones. In case it hurts innocent civilians, who will be responsible? The builder of the drone, AI programmer, the person who launches the drone, or?

AI at the Crossroads in Human History

Chapter **8**

8.1 AI at the Crossroads in Human History

8.1.1

[AI at the Crossroads in Human History - blog](#)

If we look back from the larger perspective of man who conquered his road through deserts, forests, desolate mountains and endless oceans, we see how in certain phases the cortex grew unexpectedly quickly. Such a major phase was when men reached a higher age, through better food and more fortunate climate conditions. But this was only a small issue, compared to its unforeseen side effects. It was that parents reached the age that they became grandparents and were charged with caring for their grandchildren.

Can you imagine the difference between, children left behind alone in caves while their parents were hunting and defending their territories, versus children taken care of by their parents' parents, who tried to convey their earlier experience to the young generation? It was in this era that human language and communication developed. In order to express episodic events, like happy- and traumatic experiences in grandparents' life, mimics and hand signs were no longer sufficient to label situations, emotions, and relationships between cause and effect.

That's why it is fair to say that the three-generation family became a pivot in human intellectual development, and as such launched the first generation of AI: The explicit dedicated attention of adults to inform the youngsters became a tool in a new evolutionary advantage: Survive better, find better mates, because .. your grandparents took the time and effort to prepare for life.

8.1.2

Question: If you extrapolate the entering of grandparents in youngsters' life to the current situation, you may easily scan the Middle Ages, where monasteries and minstrels spread intellect and awareness. You might also easily pass through the era of Enlightenment, where the book printing press and formal education became necessary for spreading the fast scientific harvests. The arrival of early AI made us aware that human expertise could become consolidated in data systems and even in deductive reasoning machines. Could you try and extrapolate what catalytic role AI systems may have in the second half of the 21ST century?

Think freely in terms of

1. supplanting the current model of transfer-oriented education?
2. allowing students to use computers in holistic / meaningful rather than rote memorization in school examinations?
3. Make learners keen on collaborative- rather than soloistic task performance?
etc.

Landscape of AI Technologies

Chapter **9**

9.1 Landscape of AI Technologies

9.1.1

Wikipedia: [Landscape of AI Technologies](#)

Video: Strategic Intuition: The Key to Innovation.



"... Modern science now understands how creative ideas happen in the human mind, and by knowing how it works, you can do it better. This new science overturns previous methods about how to get an innovative idea. For example, 99% of the old methods depend on brainstorming for idea generation. That's often incorrect. Strategic intuition, developed by William Duggan, Senior Lecturer in Business at Columbia Business School, offers instead a method that matches how the brain actually works. In this recording of our 30-minute webinar from January 31, 2020, you'll learn:

- *The science behind strategic intuition*
- *The steps that happen in your mind when you get a new idea*
- *How to transform those steps into a practical method for innovation problems of all kinds William Duggan and Allie Surina teach in the Strategic Intuition Online program.*

To learn more visit: <https://www8.gsb.columbia.edu/execed/...>

AI relies on a range of complex techniques, that urge AI engineers to get accustomed to mathematics, statistics, and other data sciences, as well as coding. Traditionally, AI has been based upon facts and rules, like in the world of 'declarative programming' like in PROLOG. Once a predicate entails to underlying propositions, a whole sequence of inferences would be resolved, and the result would be a 'true' statement about the world under suspect.

The professional 'genre' of those who contributed to AI predicate calculus was 'knowledge engineer'; it was the person who understood both the domain, its threads of reasoning and the soundness of human cognition. Crucial in this last one is the virtue of knowledge extraction; seducing the expert to openly 'confess' the shortcuts in reasoning. For instance: if a medical symptom triggers a general practitioner to believe the patient suffers from a bacterial infection, the candidate of tetracycline, would pop up. Tetracycline sold under various brand names, is an oral antibiotic in the tetracyclines family of medications, used to treat a number of infections, including acne, cholera, brucellosis, plague, malaria, and syphilis.

Common side effects include vomiting, diarrhea, rash, and loss of appetite. Other side effects include poor tooth development if used by children less than eight years of age, kidney problems, and sunburning easily. Use during pregnancy may harm the baby. It works by inhibiting protein synthesis in bacteria. The less-known side-effect of poor tooth development in youngsters became a typical 'sidekick' counter indication for youngsters as it leads to black teeth in most of the cases.

9.1.2

Can you Remember? What is the role of intuition in innovative idea generation? How can AI help in it?

9.1.3

Question: The plead against monotonous reasoning may also explain why teams of experts find even better solutions when one or two novices join the meetings and pose 'stupid' questions. Can you explain why?

Machine Learning

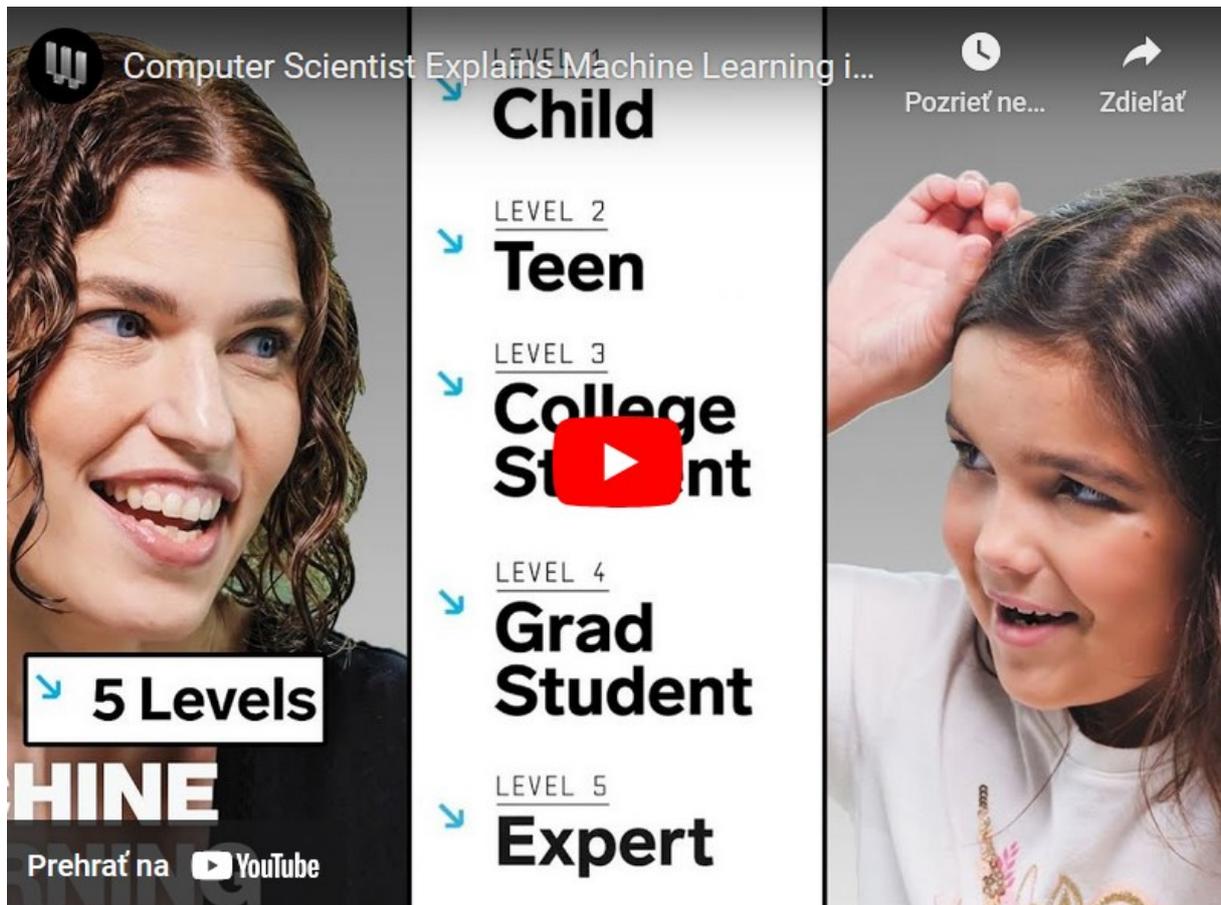
Chapter **10**

10.1 Machine Learning

10.1.1

Wikipedia: [Machine Learning](#)

Video: **Computer Scientist Explains Machine Learning in 5 Levels of Difficulty**



"WIRED has challenged computer scientist and Hidden Door cofounder and CEO Hilary Mason to explain machine learning to 5 different people; a child, teen, a college student, a grad student and an expert."

Natural language processing, facial recognition, and self-driving cars; they are all enabled by advances in computational approaches. Machine learning (ML) analyses large amounts of data to identify patterns and build a model which is then used to predict future values. Rather than calling its mechanism 'algorithmic', machine learning is now called 'learning'.

Supervised learning involves labelled data: 'Metadata'. Metadata is contextual embedding that helps learning to switch between wide-angle generalization versus narrow-angle context specificity. One should be aware that 'Machine Learning' is

not really 'learning' as we mean in a day-to-day context; the inferencing engine does not become smarter after many instances; it is just the number of cases that makes the predictive power higher; the 'learning engine' maintains the same algorithm.

The next cascade of intelligence should be remembered during the remainder of this course module:

1. Artificial Intelligence,
2. Machine Learning,
3. Neural Networks and
4. Deep Learning.

The criterion for increasing the ladder of sophistication is the amount of needed supervision; Deep learning boasts upon a higher level of autonomous learning. In both supervised and unsupervised learning, the model derived from the data is fixed, and if the data change, again the whole sequence of actions needs to be repeated again.

AI products that you see today need a continuous monitoring and correction by human experts all the time. Up to today, machine learning cannot survive without human assistance, even if the semantic context is densely described and vectorized. The Gödel 'Incompleteness Syndrome' strikes again; We always need external worlds in order to legitimize subsystems to work errorless.

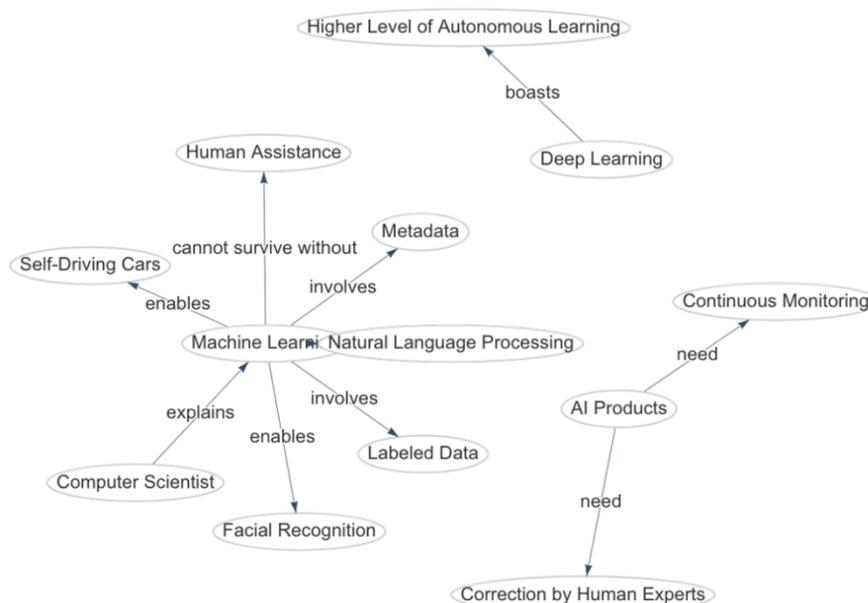


Figure: Machine Learning as Source for new A.I. Applications

 10.1.2

Can you Remember? What is the cascade in machine intelligence? Why is 'unsupervised machine learning' so important?

 10.1.3

Question: In the paragraph above, the term machine 'learning' is questioned. What would be your criterion for calling AI as 'learning' process?

AI Taxonomies

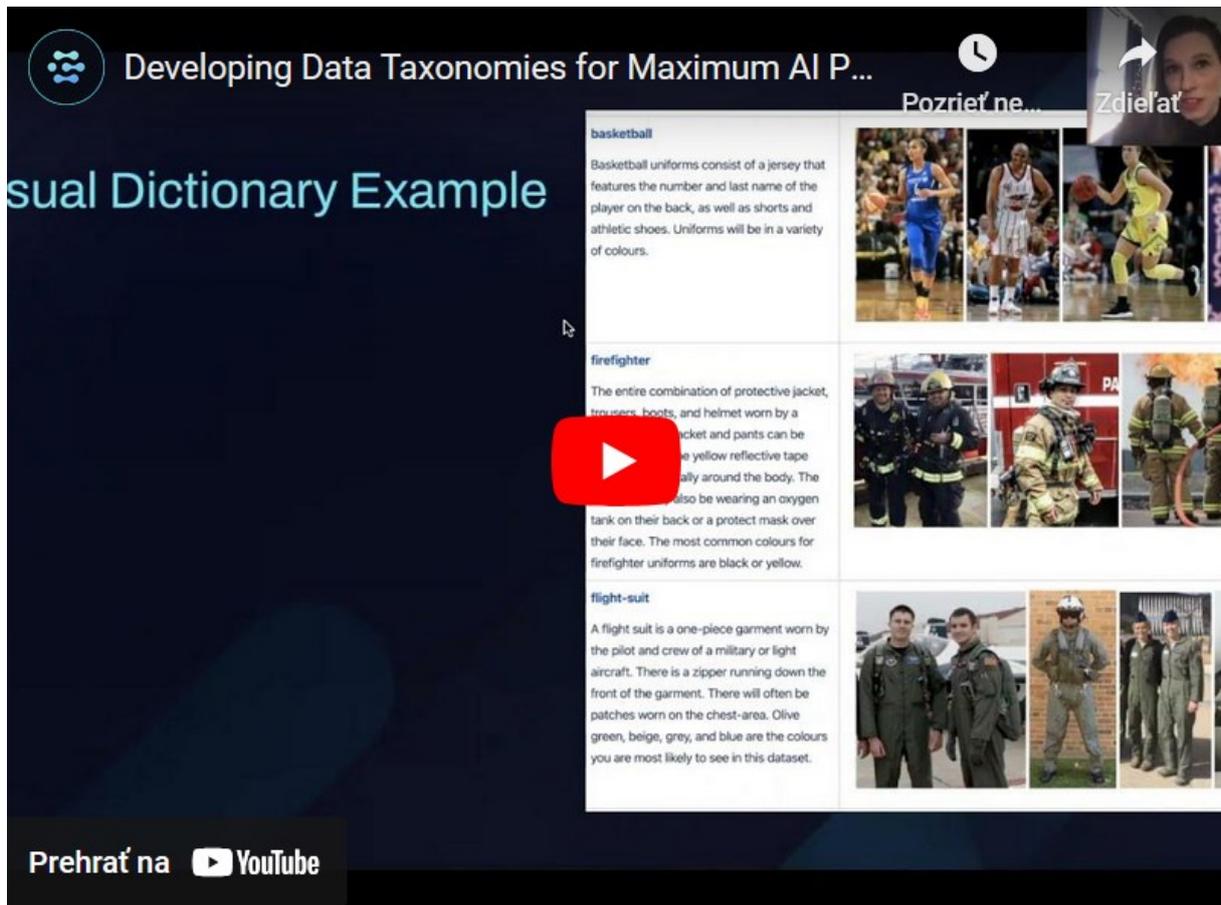
Chapter **11**

11.1 AI taxonomies

11.1.1

Wikipedia: [AI Taxonomies](#)

Video: **Developing Data Taxonomies for Maximum AI Powered Experiences.**



"Learn how to define what your product does and how it functions by defining the data taxonomy of your model. When we feed AI models unstructured data, we generally expect them to return words, and your taxonomy is their vocabulary. - Building a visually distinct taxonomy - Data sourcing to match your modelling use case - Improving model performance iteratively Learn more at: <https://www.clarifai.com/>"

Regardless of AI's need for hybrid complementary mechanisms, we see a recurring hierarchy in dependency among AI technologies as below:

- Natural language processing (NLP): The use of AI to interpret texts, based upon semantic analysis and generate texts.
- Speech recognition: The interpretation of spoken words during professional discourse.
- Image recognition and processing: Facial recognition for authorizing legal transactions, etc.

- Autonomous agents: AI in gaming, business (stock market decisions) and military actions.
- Affect detection: AI to analyse human behaviour, gaze sequences and body movements.
- Data mining: Medical diagnoses, smart city traffic optimization and tracing criminal actions.
- Artificial creativity for artistic design like generative music, story assembly, etc.

11.1.2

Table: AI Technologies for Application Strands

Technology	Details	Main AI Techniques	Development	Examples
Natural language processing (NLP)	AI to automatically generate texts (as in auto-journalism), and interpret texts, including semantic analysis (as used in legal services and translation). Machine learning (especially deep learning), regression, and K-means.	NLP, speech recognition, and image recognition have all achieved accuracy in excess of 90%. However, some researchers argue that, even with more data and faster processors, this will not be much improved until a new AI paradigm is developed.	Otter12	
Speech recognition	NLP applied to spoken words, including smartphones, personal assistants, and conversational bots in banking services.		Machine learning, especially a deep learning recurrent neural network approach called long short-term memory (LSTM).	Alibaba Cloud13
Image recognition and processing/td>	Includes facial recognition (e.g., for e-passports); handwriting recognition (e.g., for automated postal sorting); image manipulation		Machine learning, especially deep learning convolutional neural networks.	Google Lens14

	(e.g. for deep-fakes); and autonomous vehicles./td>			
Autonomous agents	Includes computer game avatars, malicious software bots, virtual companions, smart robots, and autonomous warfare.	GOF AI and machine learning (for example, deep learning self-organizing neural networks, evolutionary learning, and reinforcement learning).	Research efforts are focusing on emergent intelligence, coordinated activity, situatedness, and physical embodiment, inspired by simpler forms of biological life.	Woebot15
Affect detection	Includes text, behaviour and facial sentiment analyses.	Bayesian networks and machine learning, especially deep learning.	Multiple products are being developed globally; however, their use is often controversial.	Affectiva16
Data mining for prediction	Includes financial predictions, fraud detection, medical diagnoses, weather forecasting, business processes and smart cities.	Machine learning (especially supervised and deep learning), Bayes networks and support vector machines.	Data mining applications are growing exponentially, from predicting shopping purchases to interpreting noisy electroencephalography (EEG) signals.	Research project17
Artificial creativity	Includes systems that can create new photographs, music, artwork, or stories.	Generative adversarial networks (GANs), a type of deep learning involving two neural networks pitted against each other. Autoregressive language models that use deep learning to produce human-like text.	GANs are at the cutting edge of AI, such that future applications are only slowly becoming evident. An autoregressive language model known as GPT-3 can produce impressive human-like text. However, despite appearances, the system does not understand the text that it outputs.18	This Person Does Not Exist11 GPT-3 (Brown et al., 2020)

11.1.3

Most of the AI proponents intended to compete with human expert performances. In fact, most of the results of Table show a mediocre achievement;

First of all, because each of them restricts to task domains with a high degree of 'tractability' (numerical solvability rather than qualitative reasoning).

Secondly, as its performance limits to narrow reasoning across a small number of variables.

Thirdly, the narrative conversation with the human user is still rather primitive and needs a high human dedication.

Nevertheless, its contribution to the growing reputation of AI has been considerable. In more popular term, we can say that AI systems talk to the human user in the language of a two-year old toddler. AI's promised potential all too often remains frustratingly beyond the horizon (Lucas, 2018). It has even been suggested that progress in AI is soon to plateau (Marcus and Davis, 2019). For example, autonomous vehicles safely navigating the streets of Palermo or Delhi remain some decades away, while image recognition apps are still easily fooled (Mitchell, 2019).

11.1.4

Can you Remember? Which of the seven technologies on the ladder from basic to advanced AI technologies can you mention 'by heart'?

11.1.5

Question: The last row in the Table 3 above lists 'Artificial Creativity'. The table suggests that this is the most demanding AI activity. Do you agree? Secondly: would you AI-driven random art label as 'Artificial Creativity' as well?

Conversational Style of AI Systems

Chapter **12**

12.1 Conversational Style of AI Systems

12.1.1

Wikipedia: [Conversational Style of AI Systems](#)

TEDx: **Why conversational AI is taking over our world**



“Dr. Jason Mars is co-founder of Clinc and professor of Computer Science at the University of Michigan. He has showcased his talents in the natural processing field by becoming one of the leading entrepreneurs in conversational AI. His mission is to help solve complex, real-world issues through AI that learns on the go, empowering enterprises to use revolutionary AI to improve experience for their customers. Jason not only has an impact locally but is recognized globally as a top innovator. He continues to impact the technology industry and academia by pushing the boundaries of AI as we know it. Jason Mars, cofounder of Clinc and professor of Computer Science at the University of Michigan, has showcased his talents in the natural language processing field by becoming one of the leading entrepreneurs in conversational AI. His mission is to help solve complex, real-world issues through artificial intelligence that learns on the go. This talk was given at a TEDx event using the TED conference format but independently organized by a local community.”

Though AI evolved from the ambition to let machines think and act, its outcome has not convinced too much, except its capacity for pattern detection and statistical reasoning. Its mechanism in the latter one is multivariate, multi-dimensional analyses and inductive reasoning like in Bayesian approaches. Weak performances of AI are in autonomous learning, interpreting world-knowledge common senses and issues of morality and ethics. This antagonism has been coined as Moravec's paradox: For AI it seems relatively easy to let computers compete with adult gaming tasks like chess or GO; In terms of maneuvering through a garden, AI still looks primitive compared with a youngster of one or two years old. (Moravec, 1988, p. 15).

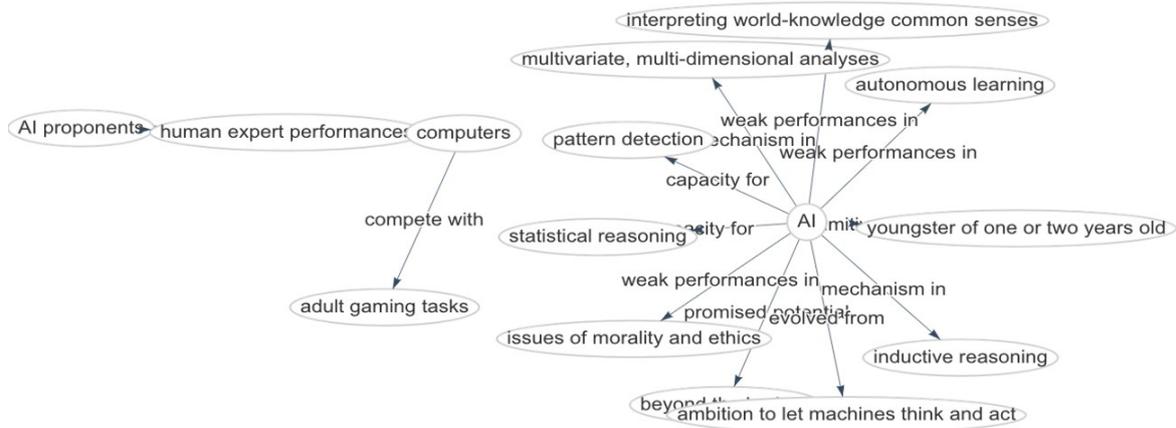


Figure: A.I. Mitigating amongst Spheres of Interest

 12.1.2

Can you Remember? Which intellectual tasks is AI relatively good in, and which prove to be much more difficult than we thought before?

 12.1.3

Question: 'Manoeuvring through a garden' is labelled by Moravec as not trivial at all. For a 2 year youngster this task is estimated as relatively easy. What do you see as 'algorithmically' complex in 'Manoeuvring through a garden'?

AI in the Context of Human Sensing, Actions and Performance

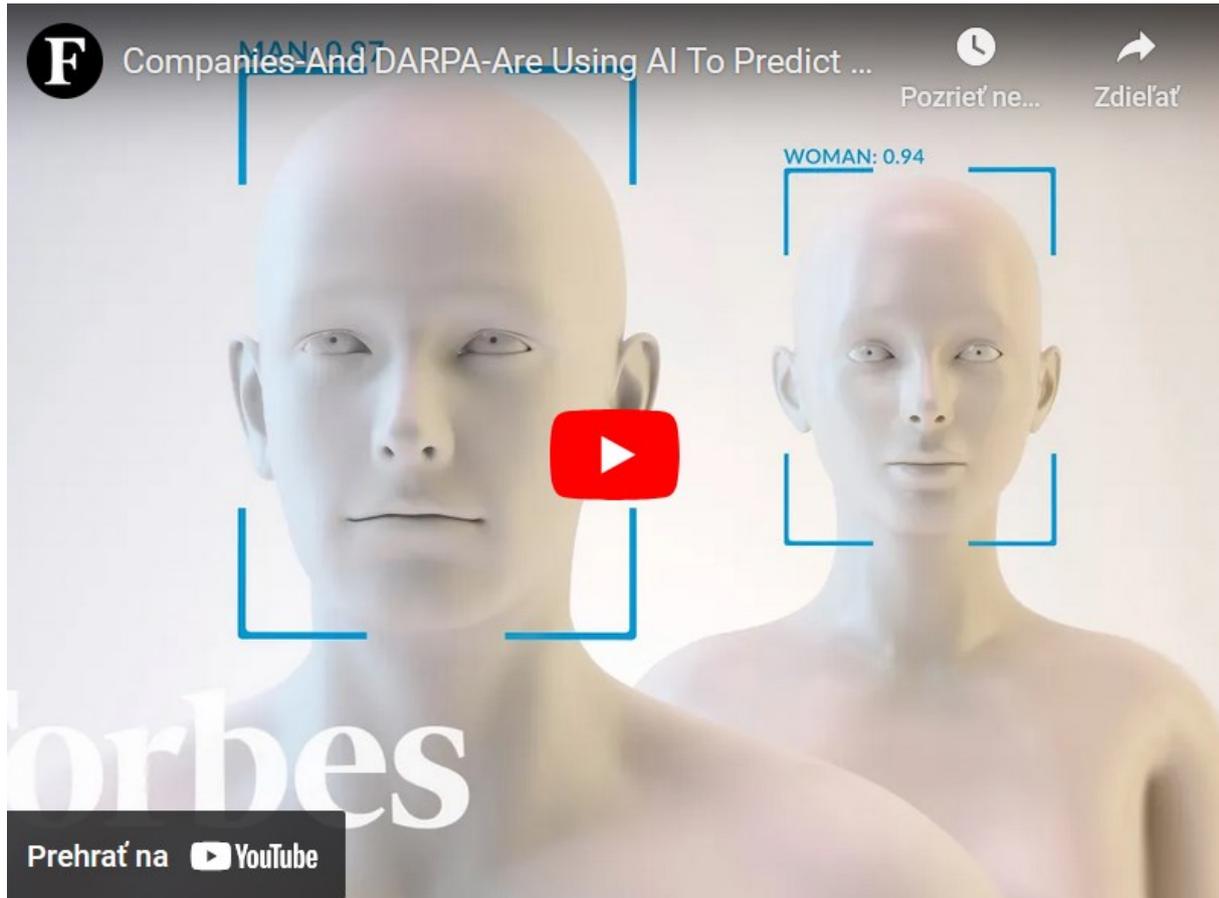
Chapter **13**

13.1 AI in the Context of Human Sensing, Actions and Performance

13.1.1

Wikipedia: [AI in the Context of Human Sensing, Actions and Performance](#)

Video: **Companies-And DARPA-Are Using AI To Predict Human Emotion.**



“ The Pentagon's research arm has pumped \$1 million into a contract to build an AI tool meant to decode and predict the emotions of allies and enemies. It even wants the AI app to advise generals on major military decisions. DARPA's backing is the starting pistol for a race with the government and start-ups to use AI to predict emotions but the science behind it is deeply controversial. Some say it's entirely unproven, making military applications that much riskier. The previously unreported work is being carried out under a DARPA project dubbed PRIDE, short for the Prediction and Recognition of Intent, Decision and Emotion. The aim is to create an AI that can understand and predict reactions of a group, rather than an individual, and then offer guidance on what to do next. Think of a military leader who wants to know how a political faction or a whole country would react should he or she take an aggressive action against their leader. In PRIDE, the emotion detection is not for an individual. It's more as a collective group and even at a national level,” says Dr. Kalyan

Gupta, president and founder of Knexus. “To think about, you know, whether a nation state is either angry or agitated.”

And it’s no small fry initiative; the plan is for PRIDE to provide recommendations for “international courses of action,” according to a contract description. Whilst DARPA’s project is largely looking at sentiment elicited from text and information posted online, a handful of start-ups, from the U.K. to Silicon Valley claim they can both understand what people are feeling and how they will feel in the future by looking at their face. In the Farringdon, London, offices of Element Human, 36-year-old founder Matt Celuszak grandly claims such emotion detection is about to cause a “shift change in how people live their lives and where humanity is evolving.” His company works with clients to hone the quality of their video ads by showing them to a small audience and having algorithms look for signs of emotion, whether that’s mild amusement or abject terror. It’s been operating largely under stealth, until now, though it’s been testing its tech with various major publishers, from CNN and Time Inc to the BBC.”

 13.1.2

In the last 50 years, the criterion for Artificial Intelligence, artificial intelligence has been constantly raised. So that some also define it as 'something that the computer cannot yet do'. Nowadays, many people find a machine intelligent if it has taught itself something, but that is not such an interesting definition for the user. A home CV thermostat is normally not seen as artificial intelligence; It is a one-to-one rule that makes the decision. More sophisticated is a thermostat that tries to estimate if the house inhabitants ‘feel cold’ or not. One step further is that the thermostat learns for the persons’ ‘taste’ and even make them learn to accept a lower temperature during years of empirical interactions. In order to distinguish between applying straightforward and self-learning, big data are needed to consider the many variables that decide humans ‘feeling of temperature’.

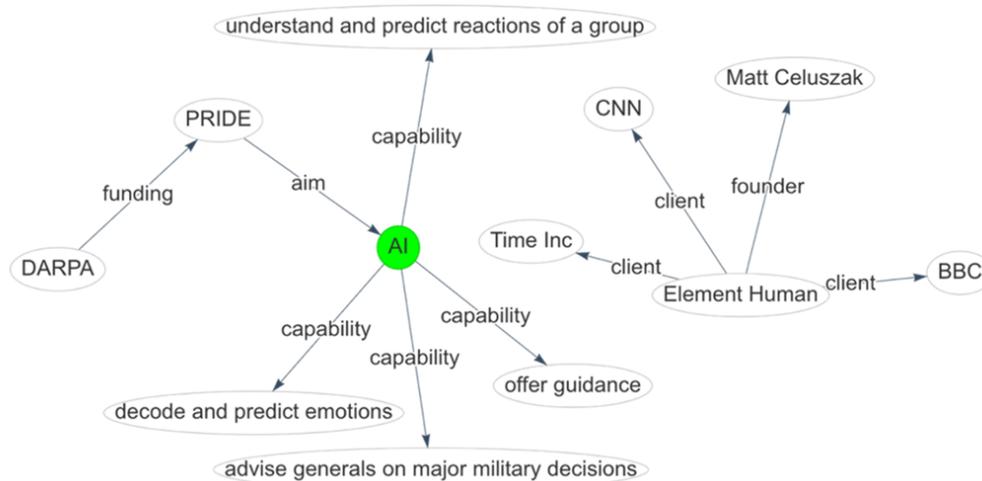


Figure: A.I. Functionalities for Faculties in Life

 13.1.3

Can you Remember? What was given as the goal for Darpa's PRIDE system? What are its easier and more complex elements?

 13.1.4

Question: The intelligence of switching on/off the heating based on actual temperature is seen as relatively easy. What factors should be taken into account if the criterion is that persons in that room should really feel a 'comfortable temperature'?

AI's Impact on Human Employment

Chapter **14**

14.1 AI's Impact on Human Employment

14.1.1

Wikipedia: [AI's Impact on Human Employment](#)

TEDx: **Preparing for a future with Artificial Intelligence**



"It's often said that history repeats itself. Many times in the course of our history, new technologies have wiped out entire workforces. For upcoming generations, the rise of artificial intelligence represents the next great solution and the next great hurdle. Robin believes how we respond to this challenge, could be our defining moment as a species. Robin Winsor is a technology leader and entrepreneur who has led several organizations from start-up to international success. Before joining Cybera as President and CEO in 2010, Winsor invented and developed the world's first direct digital x-ray system, and holds multiple patents in the medical, well-logging and seismic industries. He is a past recipient of a Manning Innovation Award, the Ernst & Young Entrepreneur of the Year Award, and the Queen's Diamond Jubilee Medal for outstanding service to Canada. He is a staunch advocate for transparency and lower cost information sharing through advancements in technology."

AI has been considered as catalytic to the transition from Industry 3.0 to Industry 4.0. The main challenge to understand is that human expertise seems to escape from the pure intellectual cascade as we tend to deliver via technical training for industry. Industry 4.0 technologies include 3D printing, autonomous vehicles, biotechnology, nanotechnology, quantum computing, robotics, and the Internet of Things. AI is all over current workplaces. The rough estimation is that in 2030, about 30% of the workplaces can be automatized. The jobs that AI can replicate and replace are those that require recently evolved skills like logic and algebra. They tend to be middle-income jobs. Conversely, the jobs that AI cannot easily replicate are those that rely on the deeply developed skills such as mobility and perception. They are usually lower-income jobs. As a result, we can say that AI is displacing middle-income jobs and will keep most lower-income jobs untouched. In 2017, the Chinese government announced its Next Generation of Artificial Intelligence Development Plan (Government of the People's Republic of China, 2017). Again, it focused on the spectrum of theoretical and practical AI approaches, including big-databased intelligence, cross-media intelligence, human-machine hybrid augmented intelligence, collective intelligence, autonomous intelligence, advanced machine learning, brain-inspired intelligence, and quantum intelligence. Both plans highlight the potential of seamless interactions between humans and AI systems and focus on the potential social and economic benefits of AI, while minimizing its negative impact.

14.1.2

Can you Remember? What percentage of the jobs have been estimated to disappear because of AI?

14.1.3

Question: One may assume that AI supplants the high-paid jobs earlier than the low-paid jobs. Do you see exceptions to this assumption?

Social Aspects of Artificial Intelligence

Chapter 15

15.1 Social Aspects of Artificial Intelligence

15.1.1

Wikipedia: [Social Aspects of Artificial Intelligence](#)

Video: **AI: Impact on Society**



" AI and Machine Learning are changing the entire world and having a huge impact on our society. Everything from food production and energy efficiency to justice and economics are being transformed by this technology. It's important that you understand how AI works so that you can have a hand in how it is being used. Start learning at <http://code.org/>"

AI technologies entail to value disputes like data ownership and the mutual agreement between students and institutes to reuse instructional data in the wider context of marketing, insurance and loan conditions. The risk is that algorithmic biases potentially undermine basic human rights. So far no research and no guidelines and policy reports have been created for ethical aspects of IA in education. (Holmes et al., 2018b, p. 552). The term 'dataveillance' has been coined (Lupton and Williamson, 2017): Who owns and who may use certain data? Learners are vulnerable to having their personal data misused or compromised, especially given that less

than 30% of countries across the world (excluding Europe) have comprehensive data protection laws in place.

Algorithms are playing an increasingly widespread role in society, automating a wide range of tasks ranging from decisions that impact whether someone gets a job to how long someone should remain in prison. However, people are increasingly recognizing that algorithms are not as neutral as they are often presented; and that, for example, they can automate biases with varying degrees of negative consequences for individuals (Hume, 2017).

Any biased analysis might impact negatively on the human rights of individual students (in terms of their gender, age, race, socio-economic status, income inequality, and so on). However, these particular ethical concerns, centred on data and bias, are the 'known unknowns' and are the subject of much discussion in mainstream AI. There are suggestions that leading technology companies' interest in 'ethics washing' is growing, in an attempt to avoid national or international regulation (Hao, 2019). We must also consider the 'unknown unknowns', those ethical issues raised by the interaction of AI and education that have yet to be identified.

15.1.2

Can you Remember? What is meant by dataveillance? and Ethics washing?

15.1.3

Questions:

- What criteria should be considered in defining and continuously updating the ethical boundaries of the collection and use of learners' data?
- How might schools, students, and teachers opt out from, or challenge, their representation in large datasets?
- What are the ethical implications of not being able to easily interrogate how AI makes decisions (using multi-level neural networks)?
- What are the ethical obligations of private organizations (product developers) and public authorities (schools and universities involved in research)?
- How does the transient nature of students' interests and emotions as well as the complexity of the learning process impact on the interpretation of data and ethics of AI applied in educational contexts?
- What pedagogical approaches are ethically warranted?

Demystification of AI

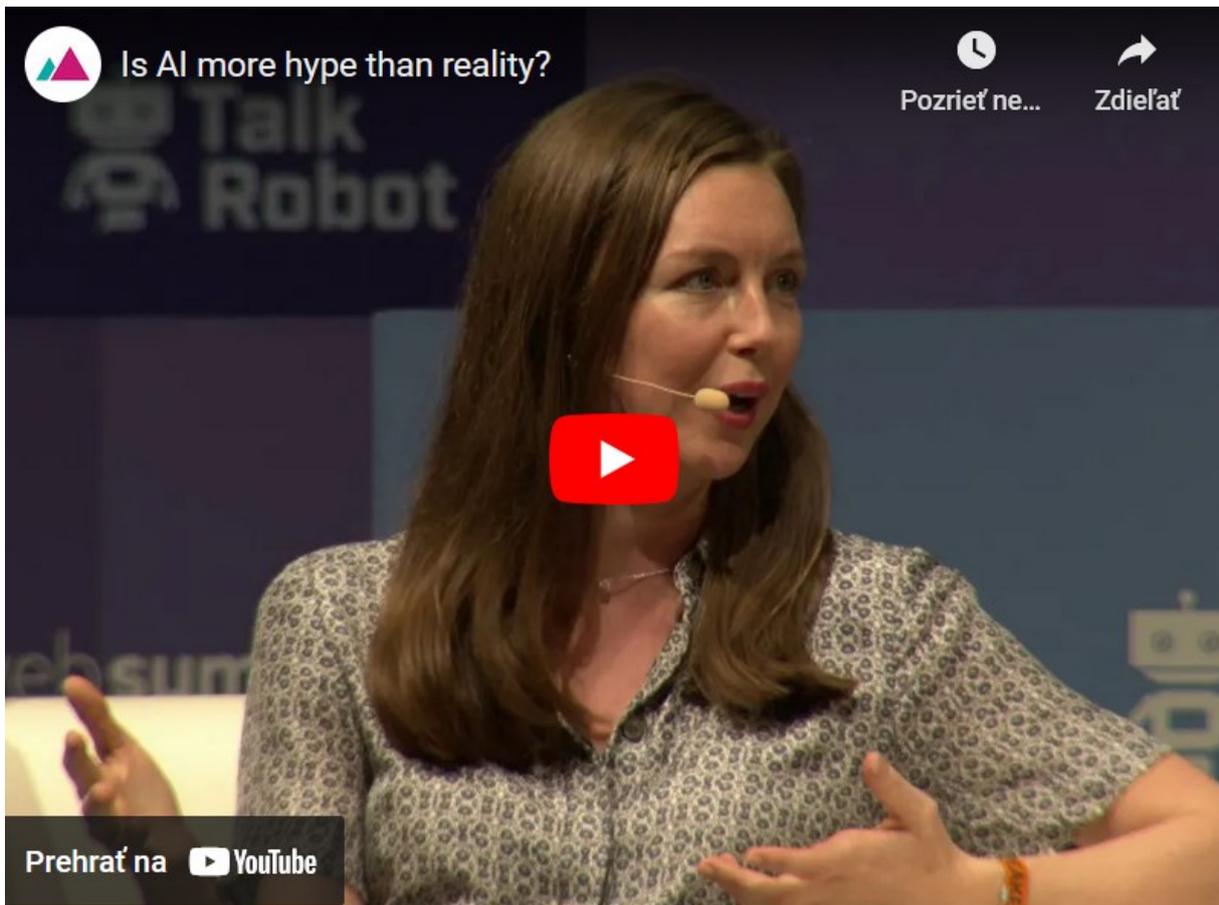
Chapter **16**

16.1 Demystification of AI

16.1.1

[Demystification of AI - link](#)

Video: **Demystifying Artificial Intelligence**



" OpenAI's Greg Brockman, M12's Nagraj Kashyap, Darktrace's Emily Orton and Bloomberg's Jeremy Kahn at Web Summit 2018. Artificial intelligence is being hailed and feared, sometimes in equal measure, as the ultimate gamechanger. Onlookers wonder whether AI will deprive us of jobs, change how we run businesses or, on the flipside, fundamentally improve our quality of life. But is the reality that AI has actually been overhyped and is now severely under-delivering? Wish you were here?"

During this course you will find a wide spectrum of techniques that underly artificial intelligence. Many of you will already think on how AI will penetrate your next job conditions: Is AI already taking a lead in smart application programs like decision taking, stock market transactions, fault finding, technical design, etc. etc. For this reason, we would like to refer to a study by [Pega](#): 'Combining artificial intelligence with human ethics for better engagement[1]'. Similarly, if we look to 'The potential for

artificial intelligence in healthcare' it is obvious that in the near future, no application direction will escape from A.I. supportive technologies. Before going to the question what ethical aspects emerge when allowing artificial intelligence to enter societies and human communities, some preliminary questions might come to your mind.

[\[1\] AI and Empathy: Combining artificial intelligence with human ethics for better engagement.](#)

 16.1.2

Can you Remember? At what points is AI hailed and feared?

 16.1.3

Question: What ethical dilemmas do you see when AI would be included in decision taken by the city council?

The Relation Between AI and Empathy

Chapter **17**

17.1 The Relation between AI and Empathy

17.1.1

[The Relation between AI and Empathy - link](#)

Video: **Empathetic AI**



"AI technologies can drive unprecedented business benefits: making intelligent decisions at scale, monetizing huge volumes of data, and jump-starting revenues – all while delighting customers. But as AI's scope increases, it becomes increasingly difficult to ensure that it will behave ethically. Building on the theme of trusted AI, Dr. Rob Walker from Pega will explore how intelligent software can be used to empower empathy within a company's engagement program, aiding the development of more sustainable, mutually beneficial customer relationships. He'll demonstrate how an organization with the commitment to constantly evolve, and the discipline to enforce ethical guidelines, can achieve tremendous results using a more humane, AI-driven engagement model."

Combining artificial intelligence with human ethics for better engagement. If you believe the robot takeover is imminent, you're not alone. Science fiction has long painted artificial intelligence (AI) overtaking and eventually enslaving humans. But

today, AI is much more benevolent, practical, useful, and already seamlessly integrated into our daily lives.

The term artificial intelligence was coined back in 1956 and today is described by the English Oxford Living Dictionary as “the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between language.” Presently, AI powers everything from personalization on websites to chatbots to automated product tagging, but even with AI delivering the types of customized, relevant experiences people demand, many consumers aren’t sold on the benefits. Perhaps the reason lies less in AI’s ability to “think,” and more in its inability to “feel.”

Cynicism for companies, strong distrust of AI. Empathy is defined as the ability to understand and share the feelings of another, or simply as “putting yourself in someone else’s shoes”. But are humans born with empathy or is it learned? Half of the audience surveyed believes human beings are born with the capacity for empathy but must learn or be taught it. That said, even with most respondents believing in our capacity for empathy, 38% don’t believe humans are the gold standard. This lack of faith extends to how respondents feel about businesses. Two-thirds (65%) of respondents don’t believe companies have customers’ best interest at heart, but 69% think organizations have a moral obligation to do what’s right for the customer (beyond what’s legally required).

These numbers aren’t surprising. In a supposed “customer centric” world, empathy is not a common corporate trait. If people believe a business doesn’t have their best interest at heart, why would they trust their technology? Even with little faith that organizations will put their needs first, the majority of respondents had an even greater distrust of AI:

- 68% trust a human more than AI to decide on bank loan approvals
- 69% say they would be more inclined to tell the truth to a person rather than AI
- 40% agree that AI has the potential to improve customer service and interactions

Perhaps some of this mistrust comes from customers not fully understanding how AI makes decisions. It seems logical to assume you could reason with a human regarding a decision on a bank loan, hoping to tap into their sense of empathy, versus a machine that has none. AI offers many opportunities for businesses to be more understanding and empathetic toward customers, but it’s up to these organizations to take control of AI, guide it, and address their customers’ concerns.

 17.1.2

Can you Remember? When was the term AI used for the first time? What is the typical trust in AI? Is it less or more than the average trust that customers have in their providers?

 17.1.3

Question: AI is often supposed to fail because of its lack of empathy. Would it be solved if AI programmers would be trained to think emphatically?

Empathy for Accountability of AI

Chapter **18**

18.1 Empathy for Accountability of AI

18.1.1

Wikipedia: [Empathy for Accountability of AI](#)

Both accountability and empathy directly relate the dimension of responsibility. In 'The role of empathy for artificial intelligence accountability' Ramya Srinivasana and Beatriz San Miguel Gonzálezb, 2022^[1]. They launch the notion that in order to consider needs of different stakeholders and thus, to put into practice accountability in Artificial Intelligence, the notion of empathy can be quite effective.

" .. Empathy is the ability to be sensitive to the needs of someone based on understanding their affective states and intentions, caring for their feelings, and socialization, which can help in addressing the social-technical challenges associated with accountability."

Systems for business compliance go increasingly fast towards empathy and use blockchain methodologies. Crucial is the duty of following laws and regulation/rules, justify the system's behaviour and face consequences accordingly. Empathy concerns the need to care for one's feelings so that pro-social attitudes become encouraged. For that reason, affective states and the intentions of stakeholders need to be 'understood': how stakeholders communicate their requirements) and how it can help AI systems to audit and report. Crucial is that empathy in A.I.-driven systems needs to be nurtured.

^[1] Ramya Srinivasan, Beatriz San Miguel González, 'The role of empathy for artificial intelligence accountability', Journal of Responsible Technology, Volume 9, 2022, 100021, ISSN 2666-6596,

18.1.2

Can you Remember? Explain the difference in accountability and empathy. Which one is easier to achieve by AI?

18.1.3

Question: The dimension accountability \Leftrightarrow empathy needs to be clarified even further. Please argue why and how human 'empathy' may have developed in evolution. Will the coming years happen the same with 'empathy' in AI systems?

 18.1.4**ADDITIONAL LECTURES**

Böhm, Steffen et al. (2022) Ethics at the Centre of Global and Local Challenges: Thoughts on the Future of Business Ethics. *Journal of Business Ethics* 180(3), 835 – 861 October 2022 ISSN 01674544 DOI 10.1007/s10551-022-05239-2

To commemorate 40 years since the founding of the Journal of Business Ethics, the editors in chief of the journal have invited the editors to provide commentaries on the future of business ethics. This essay comprises a selection of commentaries aimed at creating dialogue around the theme Ethics at the centre of global and local challenges. For much of the history of the Journal of Business Ethics, ethics was seen within the academy as a peripheral aspect of business. However, in recent years, the stakes have risen dramatically, with global and local worlds destabilized by financial crisis, climate change, internet technologies and artificial intelligence, and global health crises. The authors of these commentaries address these grand challenges by placing business ethics at their centre. What if all grand challenges were framed as grand ethical challenges? Tanusree Jain, Arno Kourula and Suhaib Riaz posit that an ethical lens allows for a humble response, in which those with greater capacity take greater responsibility but remain inclusive and cognizant of different voices and experiences. Focussing on business ethics in connection to the grand(est) challenge of environmental emergencies, Steffen Böhm introduces the deceptively simple yet radical position that business is nature, and nature is business. His quick but profound side-step from arguments against human–nature dualism to an ontological undoing of the business–nature dichotomy should have all business ethics scholars rethinking their “business and society” assumptions. Also, singularly concerned with the climate emergency, Boudewijn de Bruin posits a scenario where, 40 years from now, our field will be evaluated by its ability to have helped humanity emerge from this emergency. He contends that Milieudefensie (Friends of the Earth) v. Royal Dutch Shell illustrates how human rights take centre stage in climate change litigation, and how business ethics enters the courtroom. From a consumer ethics perspective, Deirdre Shaw, Michal Carrington and Louise Hassan argue that ecologically sustainable and socially just marketplace systems demand cultural change, a reconsideration of future interpretations of “consumer society”, a challenge to the dominant “growth logic” and stimulation of alternative ways to address our consumption needs. Still concerned with global issues, but turning attention to social inequalities, Nelarine Cornelius links the capability approach (CA) to global and corporate governance, arguing that CA will continue to lie at the foundation of human development policy, and, increasingly, CSR and corporate.

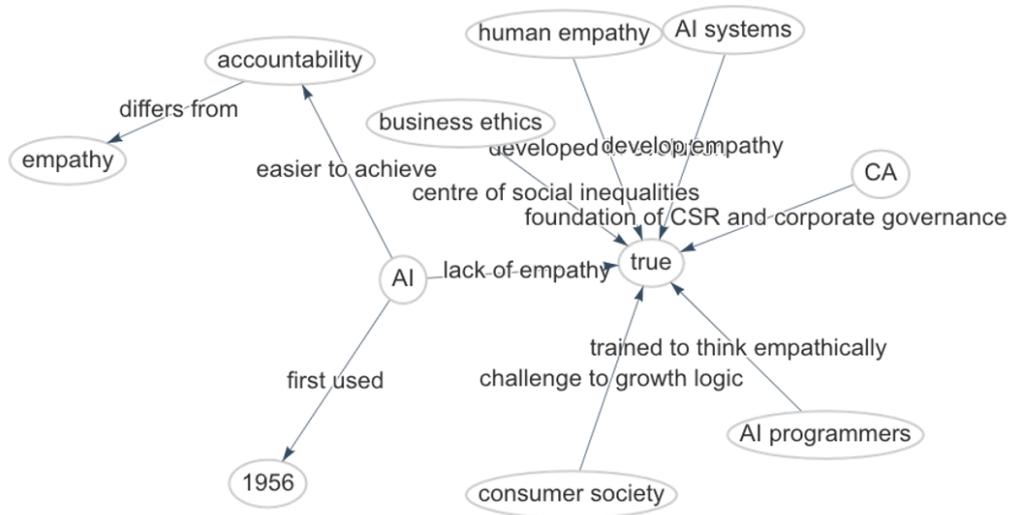


Figure: A.I. Conquering Empathy amongst Business Competition

Towards a Virtual Friend

Chapter **19**

19.1 Towards a Virtual Friend

19.1.1

Meet my friend: [Towards a Virtual Friend \(VF\)](#)

Video: **How the Internet Is Changing Friendship**



“ No matter where in the world your friends are, they’re always a text, email, or Facebook message away from reconnecting. But does that mean that we’re keeping friendships alive past their natural expiration date, or are virtual connections actually making friendships stronger? For more, read Julie Beck’s articles, “How Friendships Change in Adulthood” and “Disposable Friendships in a Mobile World”. <https://www.theatlantic.com/health/ar...> Authors: Julie Beck, Nicolas Pollock, Jeremy Raff”

Whether an andromorphic robot can finally become ‘our friend’ is the question that has been addressed by the Nobel Prize Winner Kazuo Ishiguro, as he conceived the novel ‘[Klara and the Sun](#)’. Its summary and announcement can be seen via this version on YouTube



and further in his interviews like in the following videos



"... From her place in the store, Klara, an Artificial Friend with outstanding observational qualities, carefully watches the behaviour of those who come in to browse, and of those who pass on the street outside. She remains hopeful that a customer will soon choose her, but when the possibility emerges that her circumstances may change forever, Klara is warned not to invest too much in the promises of humans..."

19.2 Concerns on the Relation Man versus Machine

19.2.1

[Concerns on the Relation Man versus Machine - link](#)

Video: **Man versus Machine**



".... This ACE lecture traces the path of an artificial intelligence radiology application that was originally developed within Erasmus MC as academic tool, but has now found its way to daily clinical practice, enabling individual patient decision-making in dementia and related brain abnormalities. For more information please visit: ace.erasmusmc.nl/lectures Presented by: prof. dr. Wiro Niessen & dr. Harro Seelaar"

Maybe science fiction's portrayal of AI has influenced some of the respondents, as 27% cited the rise of robots and enslavement of humanity as a concern. But aside from the fears of an impending robot uprising, AI is still a topic that generates legitimate concern. One of the most common fears is that AI will replace our human workforce, and this was true for the respondents as well. In fact, 35% said they were concerned about machines taking their jobs. It's not without merit, as AI was designed to perform intelligent tasks like a human – only more efficiently. While AI is commonly used to orchestrate processes that could be done by a person, these sorts of tasks are generally extremely time-consuming and tedious for employees to perform. For example, Google Translate and Netflix recommendations are both driven by AI. Can you imagine how many people it would take to translate every language request that comes in through Google or to analyse each Netflix user's viewing habits and make relevant recommendations?

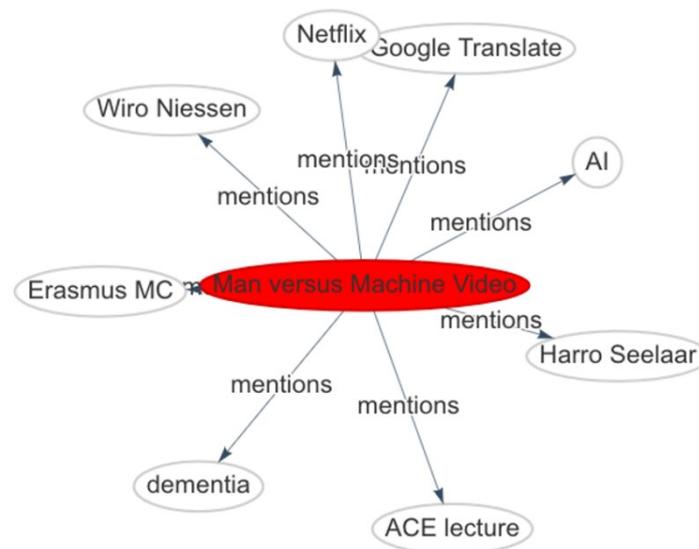


Figure: Deep Fake: Man versus Machine-generated Video

19.2.2

Can you Remember? What proportion of respondents fear to become slave of the AI application? How many percent fears to lose his/her job to AI?

19.2.3

Question: It seems that the predicate 'intelligent' jobs has at least two aspects: 1. The complexity in reasoning and thus the needed amount of computation. And 2. The amount of overall empathy and human understanding like we see as decisive for good teachers, nurses, hairdressers, etc. Could you please argue if after the arrival of AI, the empathy-job will become better-paid?

The Bias of AI

Chapter **20**

20.1 The Bias of AI

20.1.1

Think-well-group: [The Bias of AI](#)

TED Talk: **How to keep human bias out of AI**



".... AI algorithms make important decisions about you all the time -- like how much you should pay for car insurance or whether or not you get that job interview. But what happens when these machines are built with human bias coded into their systems? Technologist Kriti Sharma explores how the lack of diversity in tech is creeping into our AI, offering three ways we can start making more ethical algorithms."

54% of the of respondents believe it's possible for AI to show bias in the way it makes decisions. The same percentage felt that AI will always make decisions based on the biases of the person who created it. And 54% of respondents were sceptical about machines' ability to learn and adapt, despite science that has demonstrated this. On the other hand, it's true that AI often does have bias, from the data it collects and from bias that exists in the modern world. AI has the ability to learn, but it needs guardrails to eliminate things like bias. Businesses using AI should work to pinpoint

bias and adjust and improve AI systems using additional testing, guidelines, and controls. Perhaps the biggest concern was around AI's inability to truly emulate human interaction. About one-third (38%) of respondents don't believe AI could ever understand their preferences as well as a human being. Only 30% of respondents said they're comfortable with a business using AI to interact with them. But as AI technology becomes more sophisticated, it might not always be clear if the "person" you're interacting with is an actual human or machine.

20.1.2

Can you Remember? What is the biggest source of error in AI performance? Why would it be so difficult for AI to fully interact as a human person?

20.1.3

Question: the paragraph above argues that it is not a good idea to let ai emulate human tasks, but make them complementary to the way human approach certain tasks. for instance; a shopkeeper might look to customers' faces when they look to the price tags on the shelves. for the computer, it is quite tedious to derive emotions from human faces. the shopkeeper might better use ai for finding patterns in customers' spending behaviour and subsequently find the optimal price per certain product. could you think of another example that illustrates the trade-off between human- and artificial intelligence in business?

How to Take Control of AI with Empathy and Transparency

Chapter **21**

21.1 How to take Control of AI with Empathy and Transparency

21.1.1

Forbes: [How to take Control of AI with Empathy and Transparency](#)

Video: **Managing the risks of AI: Transparency of use**



"... In this video series, Managing the Risks of AI, Dr. Cindy Gordon, CEO and Founder of SalesChoice, and Cathy Cobey, a partner at EY, have come together to explore the rise of AI and the impacts it has on everyday business decisions. In episode one, they discuss transparency of AI use."

Empathy is not about human- versus AI; it's about using the best of what both have to offer. The future of AI-based decisioning is a combination of AI insights with human supplied ethical considerations. When many channels don't feature human agents, AI is the key to powering customer engagement and optimization across all channels. While only 9% of respondents said they were very comfortable with businesses using AI to interact with them, there's no doubt people love self-service options when they're accurate and efficient. AI can do the heavy lifting in terms decisioning. Decisioning combines machine learning, mostly to predict customer

behaviour and rules, which are supplied by humans. Both human and machine insights determine the next best action for a customer by embedding the AI-based decisioning engine inside an ethical framework. Transparency is also key to helping customers better understand how AI works. Organizations must employ transparent AI where appropriate, which means their machines can explain exactly why a decision was made. For example, when someone applies for credit card and is denied, the exact reasons are readily available.

21.1.2

Can you Remember? Which weaknesses are inherent to the current generation of AI systems?

21.1.3

Question: What examples can you remember of companies attempting the customer to understand how and why AI is needed? If you have no clue; please think of sectors like healthcare, education, car maintenance, insurance companies, etc.

Empathy: Nothing Without Action

Chapter 22

22.1 Empathy: Nothing without Action

22.1.1

Psychology today: [Empathy: Nothing without Action](#)

Video: **Breaking the Wall of AI without Empathyc**



" AI must account for human emotion to ensure that it is built to serve our goals. However, based on traditional theories of emotion, biased measures, and data rife with stereotypes, machine learning has long overlooked the bulk of what we express and perceive in real-world emotional behaviour. Hume introduces a new, data-driven scientific approach to emotion that will enable the development of truly empathic AI. With large-scale studies, Hume is collecting richer, more balanced, more diverse, unbiased emotion data by providing a degree of experimental control over the emotions experienced and expressed by participants around the world. Our team has used this data to train facial and vocal expression models that are much more accurate, nuanced, and generalizable than any that have previously been developed. These discoveries set the stage for fundamental advances in affective science and empathic technology and pave the way for AI that verifiably serves human emotion."

The next generations of AI face the challenge to show ethics and empathy more fully to the user. So, it is not just to let AI provide the correct advice or information; humans

are not satisfied by receiving correct information; the AI agent needs to convince a customer that the trace of reasoning is correct. The majority of human users expect in any dialogue a mutual initiative to understand and share a common value. Carefully listening to the partner is a first step; extrapolating its understanding is the next one.

22.1.2

Can you Remember? What is crucial in a human – AI dialogue?

22.1.3

Question: Some AI-penetrated tasks are easily adopted, as the 'victim' has no choice. For instance if you compile a CV and know from the beginning that an AI agent will sort the candidates. What is your opinion on slowly integrating AI in the citizen-government transactions? What is your recommendation for local and national authorities in this?

The Target Goals of AI

Chapter **23**

23.1 The Target Goals of AI

23.1.1

Wikipedia: [The Target Goals of AI](#)

Video: **Isaac Asimov: Three Laws of Robotics**



"... This is a clip on the Lex Clips channel that I mostly use to post video clips from the Artificial Intelligence podcast, but occasionally I post clips from other lectures by me or others. Hope you find these interesting, thought-provoking, and inspiring. If you do, please subscribe, click bell icon, and share."

If we arrive at the question on how AI relates to ethics, the "Three Laws of Robotics" (often shortened to Asimov's Laws) may come to your mind. The rules were introduced in his 1942 short story "Runaround" (included in the 1950 collection *I, Robot*).^[1]

1. First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. Second Law: A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

3. Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Asimov's laws open the doors to what we need in this course: hear our 'inner voice' and start questioning how essentially AI may work together with the long and many traditions of humans who ask themselves: "Yes the many unexplored features of AI may be possible, but are its full working out desirable as well?" In other words: "We need to develop critical thinking in order to become keen on ethical and moral implications of AI for humankind."

[1] Asimov, Isaac (1950). "Runaround". I, Robot (The Isaac Asimov Collection ed.). New York City: Doubleday. p. 40. ISBN 978-0-385-42304-5. This is an exact transcription of the laws. They also appear in the front of the book, and in both places there is no "to" in the 2nd law.

23.1.2

Can you Remember? What is the main difference between the human existential rights and Asimov's 'Laws for a Robot'?

23.1.3

Question: 'Critical Thinking' is mostly seen as attempt to fuller understand why phenomena and human conventions are as they are. If it comes to the regulation of AI, what analogy to human norms would you see as most natural?

Computational Thinking and AI

Chapter **24**

24.1 Computational Thinking and AI

24.1.1

Wikipedia: [Computational Thinking and AI](#)

Video: **Does AI make better decisions than humans? Thinking Ethics of AI.**



" Do you think that you are always making decisions by yourself? Maybe not. Your decisions may be guided by Artificial Intelligence (AI). We make many decisions every day, where to go, what to eat, what to buy, what to read, who to meet... You may listen to someone's advice, look for reviews to make your choices. But what happens if such information is biased? AI is increasingly used to assist in decision making. AI collects and analyze a large amount of data and present you, its conclusions. AI is not only used for shopping and restaurant advice, but also for more critical decision makings in our society such as medical treatment, verdict, health insurance, etc. What are the ethical concerns in using AI for decision making? What do we need to know? A new video "Does AI make better decisions than humans?" of the video series "Thinking Ethics of AI" is now available with English, French, Japanese and other languages subtitles."

One of the major sources for critical thinking is the 'inner ear'. Like sensitizing our 'outer ear' it is our conscious decision to sharpen our sense for 'good' versus 'evil', or as happens in human conflict and aggression, we may 'switch it off' for a while. As

already phrased at the start of this course module: “This course pleads for opening your mind and senses so that fundamental questions become welcomed rather than refuted.” For those of you who do not experience yet the innate eye or ear for AI’s ethical dilemmas, you may read the recent news on the big three ICT players at the moment: Google, Microsoft, and Facebook. They meet day by day societal critique on the huge (both planned and unplanned) side effects of the ocean-wide data they can use for manipulating citizens into their buying and believing behaviour. The famous Stephen Hawkins expressed “the development of full AI could spell the end of the human race. It would take off on its own and redesign itself at an ever-increasing rate. Humans who are limited by a slow biological evolution, could not compete and would be superseded”.

Further readings:

- [“To Save Everything, Click Here: The Folly of Technological Solutionism”](#) by Morosov
- [“The Every”](#) by Dave Eggers
- [“1984”](#) by George Orwell
- [“Second Life”](#)

Five years ago, Jim Stolze, co-founder of the Data Inspection Service, gave a TEDx talk about artificial intelligence in Amsterdam. Stolze looks back: 'I emphasized then that people should not be afraid that robots will take over our tasks, but that we should mainly be afraid of "the robot in ourselves". All day long, we all do stupid tasks that don't make us happy. Take something like typing numbers into spreadsheets. Those spreadsheets are better than when we did the accounting on paper. But just as spreadsheets have freed us from paper, AI is going to rid us of spreadsheets”.

24.1.2

Can you Remember? What is the goal of this course module? How does Hawkins conceive the end of the human race?

24.1.3

Question: If AI is going to rid us from spreadsheets; what will rid us from AI?

The AI Cash-Flow Formula

Chapter **25**

25.1 The AI Cash-Flow Formula

25.1.1

Blog: [The AI Cash-Flow Formula](#)

Video: **Improving cash flow forecasting with AI technology.**



"... Bank lending is the backbone to survival of at least a small-scale business. Is there a perfect rule of thumb to projections?"

A healthy cash flow is important for every company. It ensures that you can continue to pay bills and that you do not get into trouble or even go bankrupt. But how do you ensure that that cash flow is and remains healthy? The ideal cash flow formula consists of the following 5 elements, of which we will cover the first in this article.

1. Automate accounts receivable management
2. Communicate appropriately with debtors
3. Predict behaviour through artificial intelligence
4. Professionalize risk management
5. Stay in control of the flow

 25.1.2

Can you Remember? What can AI offer to cash-flow management?

 25.1.3

Question: 'Cash Flow' sounds like a rather rigid computational job. One of its five elements is obviously much more human than algorithmic. Please describe how AI would perform on this aspect?

With AI: No More Time for Stupid Tasks

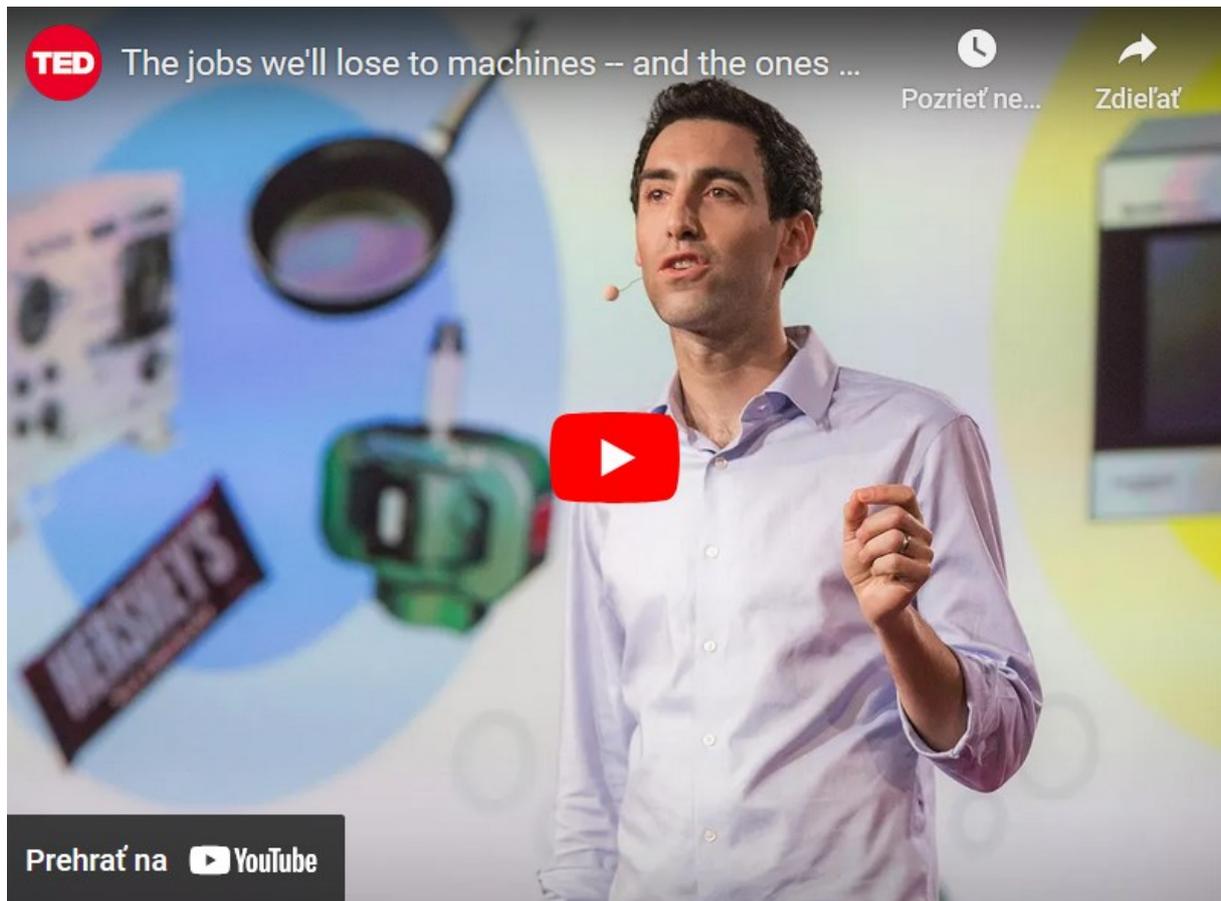
Chapter 26

26.1 With AI: No More Time for Stupid Tasks

26.1.1

AI stupidity: [With AI: No More Time for Stupid Tasks](#)

TED Talk: **The jobs we'll lose to machines -- and the ones we won't.**



".... Machine learning isn't just for simple tasks like assessing credit risk and sorting mail anymore -- today, it's capable of far more complex applications, like grading essays and diagnosing diseases. With these advances comes an uneasy question: Will a robot do your job in the future? TED-Talks are a daily video podcast of the best talks and performances from the TED Conference, where the world's leading thinkers and doers give the talk of their lives in 18 minutes (or less). Look for talks on Technology, Entertainment and Design -- plus science, business, global issues, the arts and much more."

Although the public reacted enthusiastically to Stolze's prediction, it was not really convinced yet. "Afterwards, I was even called a techno-optimist," he recalls with a laugh. 'But now we are five years further. The way people look at AI has changed quite a bit. This is partly because we are currently facing enormous staff shortages. So the less time people spend on silly tasks at work, the better. As a result, the business case for the use of an AI application is now very easy to make.'

 26.1.2

Can you Remember? What recent factor increased optimism on AI's future?

 26.1.3

Question: The arrival of web-based transactions like in the banking sector has made it easy to insert AI in banking. Still 'sitting around the table' is essential for creating trust. What situation do you remember where man-computer interaction was even more satisfactory than the human partnership?

Extra Attention to the Ethical Side

Chapter **27**

27.1 Extra Attention to the Ethical Side

27.1.1

Extra Attention to the Ethical Side

[Wikipedia intro](#)



“ With the emergence of big data, companies have increased their focus to drive automation and data-driven decision-making across their organizations with AI. While the intention is to improve business outcomes, companies are experiencing unforeseen consequences in some of their AI applications, particularly due to poor upfront research design and biased datasets. In this lightboard video, Phaedra Boinodiris with IBM, breaks down what AI ethics is and why it is so important for companies to establish a set of principals around trust and transparency when adopting AI technologies. Get started on IBM Cloud at no cost → <http://ibm.biz/get-free-acct> ”

Stolze sees that the time for experimentation is over and that more and more companies are getting serious about AI. He also sees that this new phase leads to new questions. Earlier this year, he decided to change course: he sold Aigency, an agency that builds AI solutions, and has since been active for a large part of his time within the Data Inspection Service, which he co-founded. 'We are, say, the "white

coats" that check the data for "bias" at organizations. We also subject their algorithms to stress tests in which we look at the explainability, reproducibility and honesty. There is now a great need for this extra attention to the ethical side of AI. In addition to data engineers and a lawyer, there are also two ethicists and a philosopher in our team. As the Data Inspection Service, we don't say what companies should do, but we ask questions. What algorithms do they use? And what data? That's how we go into depth.'

27.1.2

Can you Remember? What is the role of ethicists in AI teams?

27.1.3

Question: Biased data are a notorious source for failing AI. If it comes to 'big data'; what problem is often met?

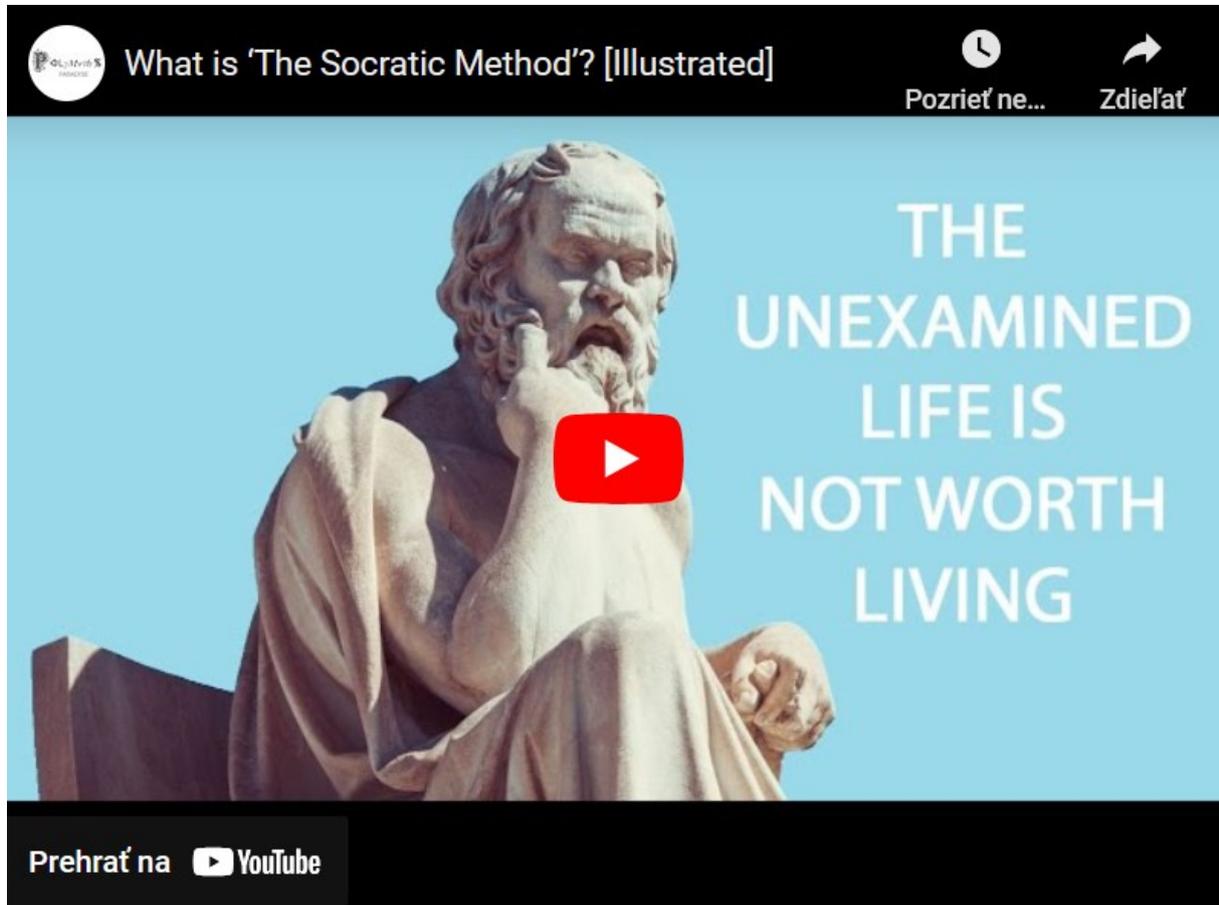
AI Socratic Method

Chapter **28**

28.1 AI Socratic Method

28.1.1

[AI Socratic Method](#)



" If you're familiar with philosophy or classics, you're bound to have seen the name 'Socrates' mentioned once or twice. Perhaps you've even seen him crop in pop culture- after all, he is very famous. But why is the so-called "Father of Western Philosophy" so important? And what can we learn from his approach to philosophy? This illustrated guide aims to answer these questions!"

The [Accenture AI Division](#) claims: " .. Although 90% of today's businesses have adopted cloud, only one third are achieving the anticipated ROI. The most advanced companies understand that while cloud sets you up with next-level computing power and access to new kinds of data in the right quantity and quality, AI is the bridge to convert that [data into business value](#). It's no surprise the entire C-suite is now involved in the AI agenda and they're asking what's next.

That's where Applied Intelligence comes in. We believe that [cloud is the enabler](#), data is the driver, and AI is the differentiator. We bring them together to help you make smarter, faster decisions that help change your organization and enable growth—at scale. And because we understand that people are central to the success of any

technology transformation, our global team of experts bring the cross-functional skills to both deliver business outcomes and facilitate cultural change – empowering your workforce to use data and [AI responsibly](#).

 28.1.2

Can you Remember? What is according to Accenture crucial in ‘technology transformation’?

 28.1.3

Question: A strong example of AI making a contribution to governance is the consequential analysis in law-making: What exceptions need to be defined before a new law can fully become rolled out. Can you find another application where AI seems to work out well?

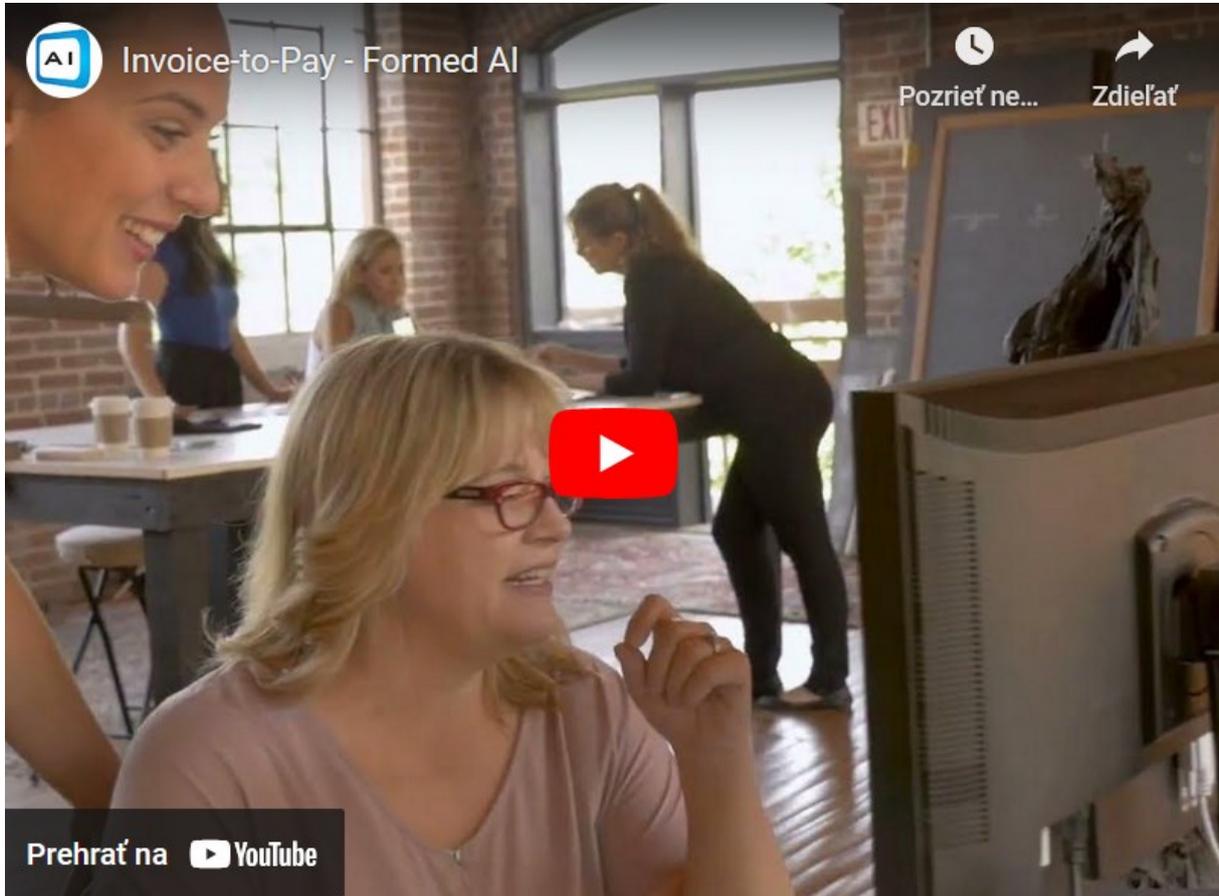
Predicting when Invoices will be Paid: Payt's AI Solution

Chapter **29**

29.1 Payt's AI solution

29.1.1

[Predicting when Invoices will be Paid: Payt's AI solution](#)

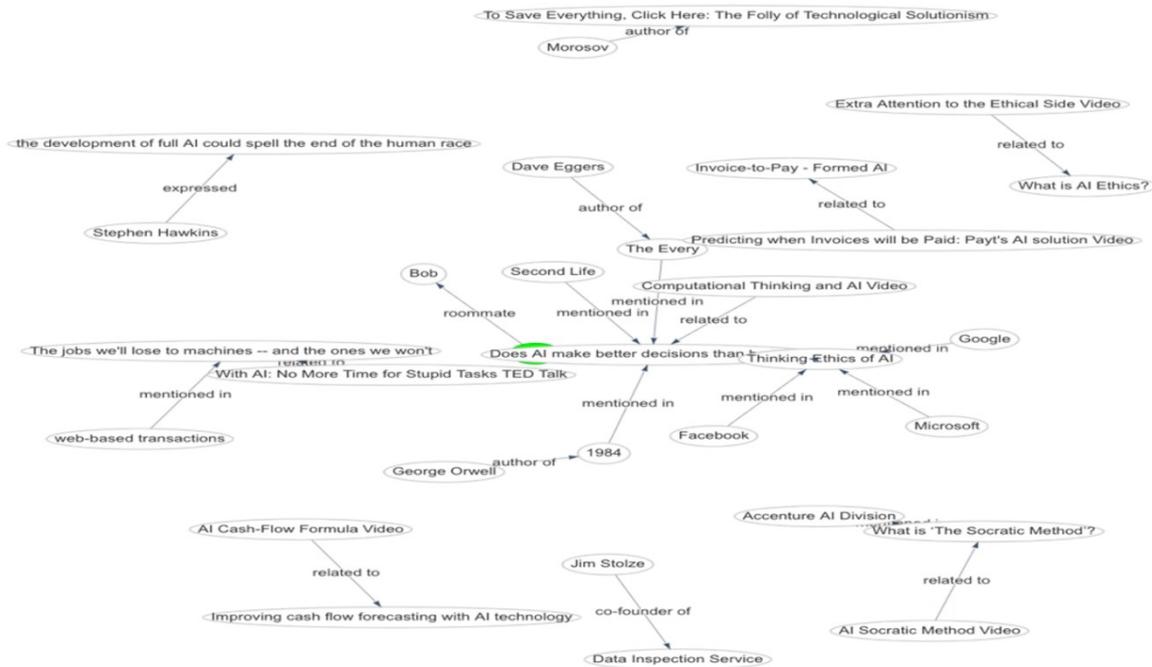


“ .. are becoming more and more sophisticated. This also applies to the quality of the data. And where it is very difficult for people to sift through thousands of financial transactions and look for things that deviate, AI applications are very good at that. So Payt decided to develop an AI application himself.”

Socially responsible choices: 'When building our AI application, we started small and simple', emphasizes Laura Baakman, software engineer at Payt. 'And we are now expanding that step by step. First of all, we look at how we can teach the algorithm to predict as accurately as possible when our customers' debtors will pay their bill. Many different factors play a role in this. Which industry is involved? Is it a private customer or a business customer? We are adding more and more parameters. In addition, there are also things that are a challenge. Adding the postal code, for example. Because is that desirable, to consider the neighborhood where someone lives? We ourselves think it is a case of doubt. It's not just about what's technically possible, but also how we can do this in a socially responsible way.'

29.1.2

Will A.I. help Humans to Make Better Decisions?



The practice Payt's AI module has been in beta with some customers since the beginning of 2022. With each invoice, employees of the accounts receivable administration see what the chance is that an invoice will be paid late and what the expected payment date is. With this information, the debtor administration can better determine their approach. For example, the announcement that a debtor is likely to pay late can ensure that that customer receives a payment reminder just a little faster than what is usual. In the long term, Payt will offer the possibility to proceed to automated follow-up actions based on the payment probability calculation.

Laura Baakman: 'However, we do this very carefully with attention to all sides of AI, including the ethical ones. The beginning is there and we look forward to making the AI module available to all our customers in the short term. One thing is certain, the possibilities are enormous.'

29.1.3

Can you Remember? What data sees Payt as disputable data in an AI-oriented customer database?

AI in the Social Context

Chapter **30**

30.1 Artificial Intelligence in the Social Context

30.1.1

Artificial Intelligence in the Social Context



“ Terah Lyons is the Founding Executive Director of the Partnership on AI, which was established to study and formulate best practices on AI technologies and advance the public’s understanding of AI. Prior to this, she served as Policy Advisor to the U.S. Chief Technology Officer at the White House Office of Science and Technology Policy (OSTP) during the Obama administration. In this video, Terah shares how after the California State Senate proposed using a machine learning-based system for pre-trial detention decisions, the Partnership investigated the plan and published a report on the minimum requirements such a system would need to meet to be fit for use. The requirements held not just for the system itself, but also for the people using it: for example, judges would need to demonstrate an understanding of the algorithm’s outputs.”

A machine is called self-learning if it acquires intelligence by teaching itself how to classify or value incoming data (patterns). In the learning phase, it is necessary that the machine can test its decision about the data (the pattern) in order to learn what a good judgment is. For known problem domains, there are three ways to do this:

1. [Supervised learning](#): The machine is trained with many patterns that have already been pre-labelled by people; for example, thousands of photos tagged with "dog" or "cat." Or many thousands of hours of video of speeches in which the machine learns to lip-read, the 'labelling' is already there in the form of the sound of the video. It is particularly costly to have datasets labelled to train machines for patterns in a new problem domain. Self-learning should therefore certainly not be associated with 'cheap learning', it is rather that it can tackle a class of problems that cannot be properly understood with classical programming.
2. [Reinforced learning](#). There are problem domains with a very easy test on classification, namely: 'do I have much or little success?' An example is learning the board game 'Go'. The rules of are simple, the strategy to win, on the other hand, is rather unfathomable and until ten years ago was considered too complicated for computers. However, with sufficient processing capacity and a cleverly set up deep neural network, it turned out to be possible to let a computer play games go against itself, after which after a few hours game patterns emerged that belong to beginners and after a day of very advanced strategies after three days the computer could compete with professional world players. (See also: <https://deepmind.com/blog/article/alphago-zero-starting-scratch>).
3. [Unsupervised learning](#). There is also a third way, in which the machine itself searches for classifying characteristics in unlabeled data. In addition to self-learning machine, we can also give a machine 'intelligence' through classical programming with many logical rules, with which the machine assesses the incoming data according to the fixed programmed logic. Self-learning machines are therefore especially suitable for classifying patterns that cannot be easily judged with classical logic. The behaviour of most industrial production processes can be translated relatively easily into classic programming rules and simulation models. But the distinction between images of a dog and a cat is not easy to make with classical programming. When we think of pattern recognition, we should not immediately think of images, it is more often about other patterns such as behavioral patterns, syndromes, traffic flows, jurisprudence, lip reading, wear and tear of infrastructure, etc.

30.1.2

Can you Remember? What are the three levels of autonomy in machine learning?

30.1.3

Question: Why is 'unsupervised' learning that essential in the future of advance AI?

30.2 The Risk of Trivial Findings in Big Data

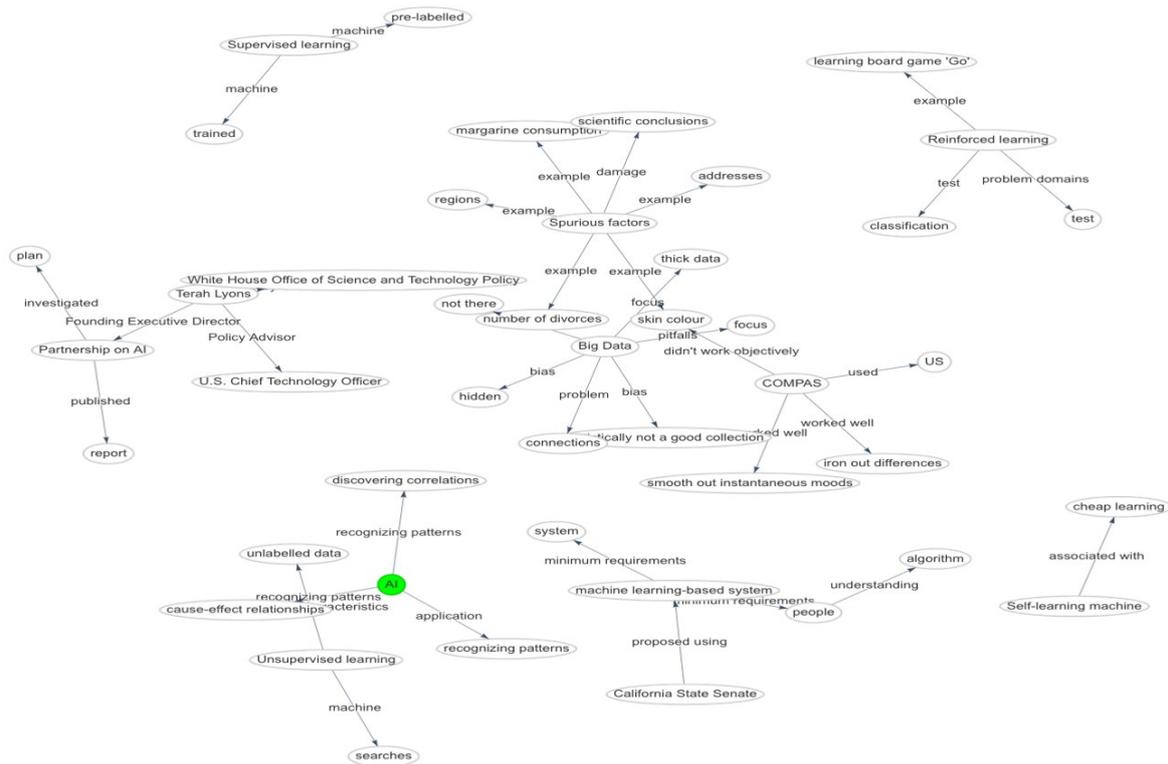
30.2.1

The Risk of Trivial Findings in Big Data



“ Why do so many companies make bad decisions, even with access to unprecedented amounts of data? With stories from Nokia to Netflix to the oracles of ancient Greece, Tricia Wang demystifies big data and identifies its pitfalls, suggesting that we focus instead on "thick data" -- precious, unquantifiable insights from actual people -- to make the right business decisions and thrive in the unknown.”

Recognizing (data) patterns is the most commonly used application of AI; it involves, among other things, discovering (learning) correlations and cause-effect relationships. The collection, analysis, interpretation, and recognition of information in large amounts of data are simply referred to as the field of big data. In large collections of data, bias is often hidden. This means that the data collection is statistically not a good collection for the goal we want to achieve. Notorious, for example, is the 'algorithm' COMPAS that is used in the US to support judges. It worked well to iron out differences between individual judges.



And it worked well to smooth out instantaneous moods of a judge; judges appear to judge more reasonably after a meal, which is the influence of the glucose level. But it didn't work objectively with respect to skin color because the data set from which it was learned was not neutral. Not counting skin color still leads to discrimination because addresses and regions also correlate with skin color. Another problem with big data is that connections are found that are not there. For example, there seems to be a strong link between margarine consumption and the number of divorces in Maine (US), if we combine the data on well-being with agricultural products.

This link may astound you <http://www.tylervigen.com/spurious-correlations>.

30.2.2

Can you Remember? What are spurious factors? Can you come up with your own example of a spurious factor and how this can damage scientific conclusions?

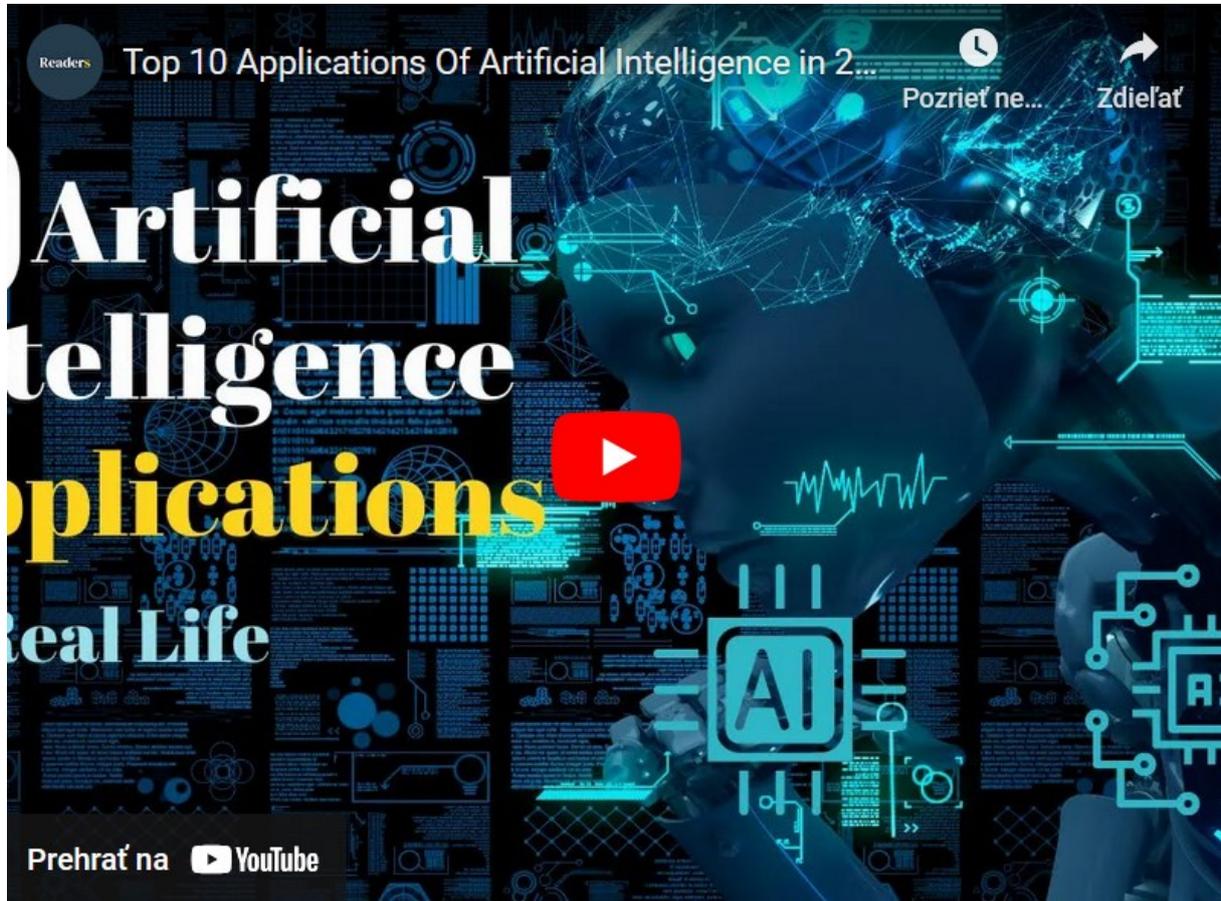
30.2.3

Question: In order to prevent from 'profiling', citizens on the basis of race, social status or sexual orientation, it seems crucial not to have these variables accessible in investigation policies. From the other side, efficiency in Law enforcement is opportune as well. What rational would you apply for finding an acceptable balance between these two?

30.3 Current Applications of Artificial Intelligence

30.3.1

Current Applications of Artificial Intelligence



“ ...Welcome Readers. In this video we brought you information about Top 10 applications of artificial Intelligence. AI is dominating our world. These applications of ai predicts how beneficial the future of ai can be. We have compiled the list of top 10 Applications of Artificial Intelligence ai. [00:00](#)-Introduction [00:55](#)-AI in Ecommerce [01:38](#)-AI in Navigation [02:11](#)-AI in Robotics [02:44](#)-AI in Human Resources [03:09](#)-AI in Health [03:45](#)-AI in Music Production [04:34](#)-AI in Gaming [05:11](#)-AI in Automobiles [05:45](#)-AI in Social Media [06:21](#)-AI in Marketing [07:18](#)- Q & A”

According to [Wikipedia](#): “ .. Google Maps is a [web mapping](#) platform and consumer application offered by [Google](#). It offers [satellite imagery](#), [aerial photography](#), street maps, 360° [interactive panoramic](#) views of streets ([Street View](#)), real-time traffic conditions, and [route planning](#) for traveling by foot, car, bike, air (in [beta](#)) and [public transportation](#). As of 2020, Google Maps was being used by over 1 billion people every month around the world.”

 30.3.2

Can you Remember? Do you agree that Google maps frees the driver from cognitive load? What other AI-driven applications for the consumer can you mention? Do you see negative side effects for the user in the long run?

 30.3.3

Question: Vehicles so far help us to migrate persons and goods. The issue of smart mobility can be seen as just a symptom of modern societies to save time. As you consult the latest news on **Metaverse**, it may become clear that virtual presence will take over large parts of physical displacements. What will become the scenario for being in a self-driving car; Is it likely that the 'dismissed driver' stays virtually present with his/her origin? Or: is it more likely that the opposite is going to happen: Stay physically at your own location, and travel virtually around to place you want to visit?

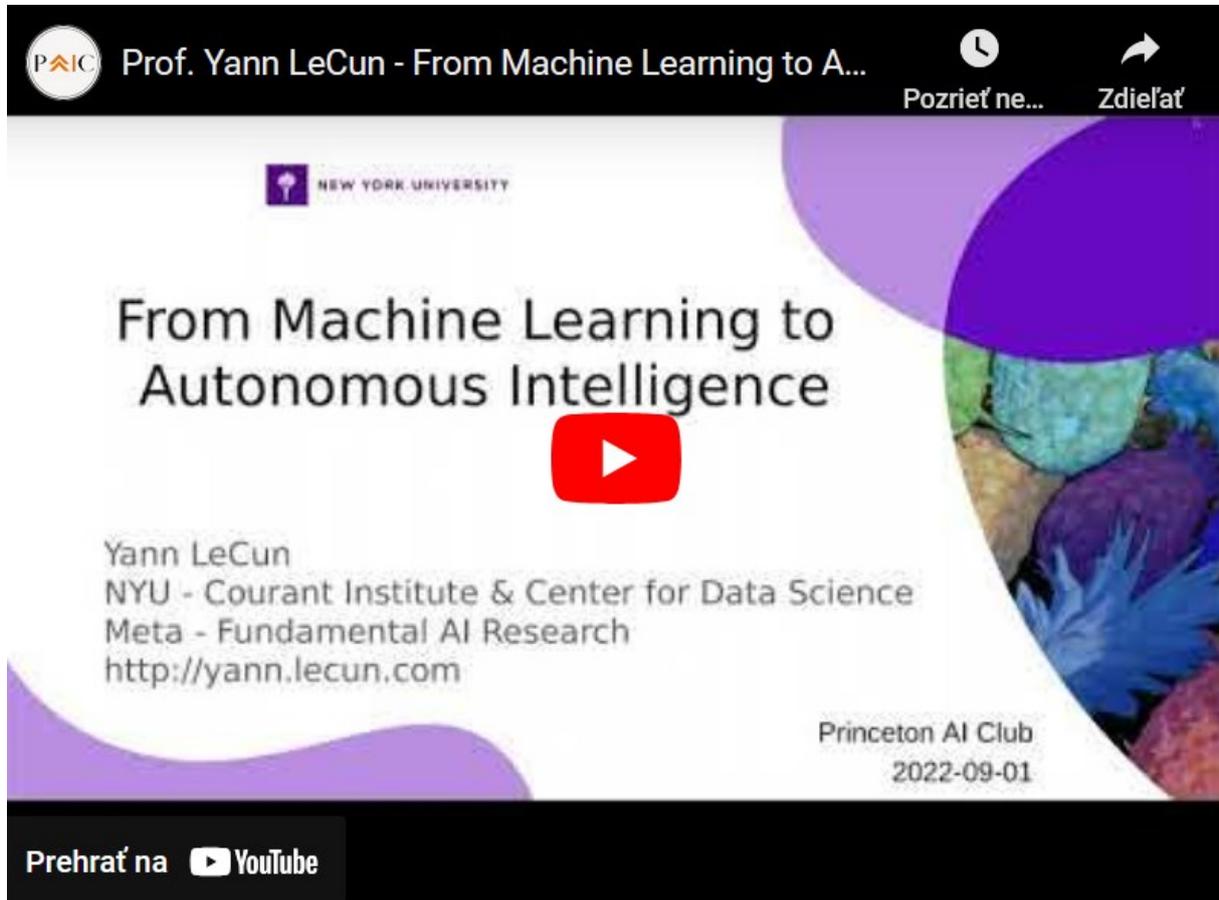
Definition of AI

Chapter **31**

31.1 Definition of AI

31.1.1

[Definition of AI](#)



“ How could machines learn as efficiently as humans and animals? How could machines learn to reason and plan? How could machines learn representations of percepts and action plans at multiple levels of abstraction, enabling them to reason, predict, and plan at multiple time horizons? I will propose a possible path towards autonomous intelligent agents, based on a new modular cognitive architecture and a somewhat new self-supervised training paradigm. The centerpiece of the proposed architecture is a configurable predictive world model that allows the agent to plan. Behaviour and learning are driven by a set of differentiable intrinsic cost functions. The world model uses a new type of energy-based model architecture called H-JEPA (Hierarchical Joint Embedding Predictive Architecture). H-JEPA learns hierarchical abstract representations of the world that are simultaneously maximally informative and maximally predictable.”

According to [Wondrium Data](#): “ .. When discussing artificial intelligence, or what many prefer to call autonomous learning machines, an important distinction must be made, which lies in the word ‘learning’. There are many autonomous machines in existence already. While these machines can operate independent of human control,

they're not, generally, adaptive. They don't learn from experience. They don't adapt to unanticipated situations. They only do what they're programmed to do."

31.1.2

Can you Remember? What is 'learning for AI machines' in the most optimal sense?

31.1.3

Question: Neural networks have proven the power of 'learning by many examples'. Human learning also uses this 'informal learning'; not a guided process otherwise than social attention, curiosity and the will to survive and care for offspring. Can you think of more examples of neural learning than distinguishing dogs and cats?

31.2 Singularity in Artificial Intelligence

31.2.1

[Singularity in Artificial Intelligence](#)



TED Talk: What happens when our computers get smarter than we are?

“ Artificial intelligence is getting smarter by leaps and bounds – within this century, research suggests, a computer AI could be as "smart" as a human being. And then, says Nick Bostrom, it will overtake us: "Machine intelligence is the last invention that humanity will ever need to make." A philosopher and technologist, Bostrom asks us to think hard about the world we're building right now, driven by thinking machines. Will our smart machines help to preserve humanity and our values – or will they have values of their own?”

According to [Wikipedia](#): “ .. The technological singularity—or simply the singularity^[1]— is a [hypothetical](#) future point in time at which technological growth becomes uncontrollable and irreversible, resulting in unforeseeable changes to human civilization.^{[2][3]} According to the most popular version of the singularity hypothesis, [I.J. Good's intelligence explosion](#) model, an upgradable [intelligent agent](#) will eventually enter a "runaway reaction" of self-improvement cycles, each new and more intelligent generation appearing more and more rapidly, causing an "explosion" in intelligence and resulting in a powerful [superintelligence](#) that qualitatively far surpasses all [human intelligence](#).^[4]”

31.2.2

Can you Remember? What in the context of Singularity is a ‘runway reaction’? Will it finally impair human intellect, or will it shift human thinking towards spiritual concerns?

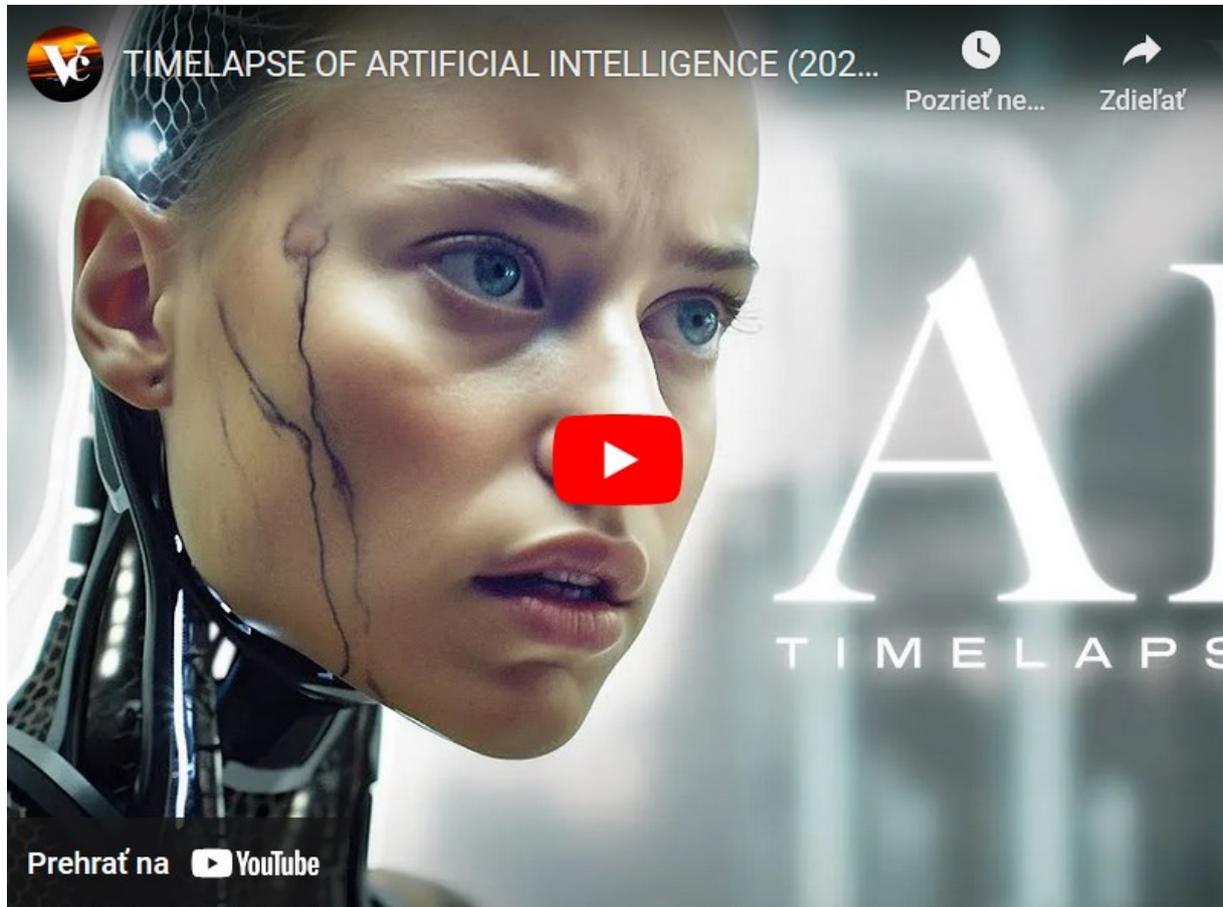
31.2.3

Question: The premise at the start of the paragraph above is that AI will help to improve itself. We already know for two decades the software that generates the layout of PCBs (printed circuit boards) and electronic traces on silicon chips. However, more recently, the AI ‘[Alpha Go](#)’ succeeded to discover the ‘rules of thumb’ for winning the enormous complex middle stage in the Go game. It is a vivid exemplar that AI is feeding itself so that we might arrive earlier at the stage of singularity. Please check the coming week which other specimens of self-propelling AI are being announced via the media.

31.3 Timeline of Artificial Intelligence

31.3.1

Timeline of Artificial Intelligence



“ A documentary and journey into the future exploring the possibilities and predictions of artificial intelligence. This timelapse of the future explores what is coming, from robots that are too fast for humans to see, to A.I. bots from Microsoft (bringing back loved ones to life). Elon Musk’s Neuralink goes from a medical and healthcare device, to helping people become superhuman – with intelligence amplification, and add-ons that connect to the brain chip. Artificial general intelligence begins to design an A.I. more powerful than itself. People begin to question if humanity has reached the technological singularity. Artificial Super Intelligence emerges from the AGI. And further into the deep future. Human consciousness becomes digitized and uploaded into a metaverse simulation. It is merged with A.I. creating hybrid consciousness – which spreads across the cosmos. Matrioshka brains and Dyson Spheres host humanity’s consciousness in a cosmic simulation network.”

Due to [Wikipedia](#): “ .. The term 'artificial intelligence' dates back to the forties of the last century and has long carried the odium 'Science Fiction' with it. The English term "science fiction" literally means “fictitious science” given by its creator [Hugo](#)

[Gernsback](#). For Gernsback, science fiction was primarily intended as a means of popularizing science and scientific discoveries, but its history shows that this idea was pursued only to a limited extent by later writers. In the genre, nowadays, images of the future with invented technological progress and scientific developments are mainly used. Critic [Darko Suvin](#) uses the following definition: "a literary genre or word construction whose conditions are the presence and interaction of alienation and cognition, and whose main instrument is an imagination framework that provides an alternative to the author's empirical environment."

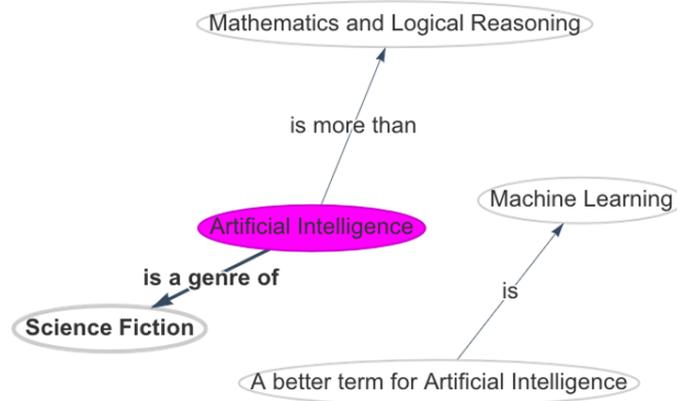
31.3.2

And in [Wikipedia](#): " .. In antiquity, the archetype of AI was loaded with myths, stories and rumors of artificial beings endowed with intelligence or consciousness by master craftsmen. The seeds of modern AI were planted by philosophers who attempted to describe the process of human thinking as the mechanical manipulation of symbols. This work culminated in the invention of the programmable digital computer in the 1940s, a machine based on the abstract essence of mathematical reasoning. This device and the ideas behind it inspired a handful of scientists to begin seriously discussing the possibility of building an electronic brain.

The field of AI research was founded at a workshop held on the campus of Dartmouth College, USA during the summer of 1956. Those who attended would become the leaders of AI research for decades. Many of them predicted that a machine as intelligent as a human being would exist in no more than a generation, and they were given millions of dollars to make this vision come true. Eventually, it became obvious that commercial developers and researchers had grossly underestimated the difficulty of the project. In 1974, in response to the criticism from James Lighthill and ongoing pressure from congress, the U.S. and British Governments stopped funding undirected research into artificial intelligence, and the difficult years that followed would later be known as an "AI winter".

Seven years later, a visionary initiative by the Japanese Government inspired governments and industry to provide AI with billions of dollars, but by the late 80s the investors became disillusioned and withdrew funding again. Investment and interest in AI boomed in the first decades of the 21st century when machine learning was successfully applied to many problems in academia and industry due to new methods, the application of powerful computer hardware, and the collection of immense data sets.

A.I. losing its Role for Science Fiction?



 31.3.3

Can you Remember? What is the relation between AI and Science Fiction? Is AI more than mathematics and logical reasoning?

 31.3.4

Question: The Wikipedia article on the origin of Artificial Intelligence mentions the emergence and decline of believing that computers could finally reach the level of human thinking. Looking back to the "AI winter" in the late seventies may make us more critical on the term 'Artificial Intelligence'. Seen the current industrial eagerness for machine learning, deep learning, data analytics and big data, we might better choose a more adequate term for AI. What would be a better term, in order to avoid common fear towards threatening singularity?

31.4 Artificial General Intelligence

31.4.1

Artificial General Intelligence



“ Danny Lange - VP of AI and ML at Unity Technologies and previously led innovative ML teams at Uber, AWS and Microsoft ABSTRACT Join this session to discuss the role of intelligence in biological evolution and learning. Danny Lange will demonstrate why a game engine is the perfect virtual biodome for AI’s evolution. Attendees will recognize how the scale and speed of simulations is changing the game of AI while learning about new developments in reinforcement [...]”

According to [Wikipedia](#): “ .. Artificial General Intelligence (AGI) is the ability of an intelligent agent to understand or learn any intellectual task that a human being can. It is a primary goal of some artificial intelligence research and a common topic in science fiction and futures studies. AGI is also called strong full or general intelligent action, although some academic sources reserve the term "strong AI" for computer programs that experience sentience or consciousness. Strong AI contrasts with weak AI (or narrow which is not intended to have general cognitive abilities; rather, weak AI is any program that is designed to solve exactly one problem. (Academic sources reserve "weak AI" for programs that do not experience consciousness or do

not have a mind in the same sense people do.) A 2020 survey identified 72 active AGI R&D projects spread across 37 countries.”

31.4.2

Can you Remember? Starting from the premise of ‘Strong AI’, should we limit AI to the ultimate tasks that human experts can solve?

31.4.3

Question: Artificial General Intelligence (AGI) can be seen as the more ambitious attempt to let computer software reach the level of consciousness and even the level of ‘wisdom’. What are your considerations to believe this high ambition; is it realistic; or is it simply beyond our cognitive capacity to think about it?

Artificial Life

Chapter **32**

32.1 Artificial Life

32.1.1

[Artificial Life](#)



“ Particle Life Simulation Primordial Soup - Evolution Conway's game of life Cellular automata and: Self organizing patterns JavaScript programming. Even if you don't have access to Ventrella's code, the main difference of this project with the other particle life projects is that it didn't implement collision detection, and this made simulating thousands of particles possible in real-time. Also, I added GUI controls to change the parameters in real-time this allows easy fine-tuning & exploration; hence, I was able to find some never-seen-before patterns emerge from some extremely simple models of relations. The code here is probably an order of magnitude simpler than any other Artificial Life codes out there because I started this code solely as an educational material for non-programmers and general audience to prove the point that complexity can arise from simplicity.”

According to [Wikipedia](#): “ .. Artificial life (often abbreviated ALife or A-Life) is a field of study wherein researchers examine systems related to natural life, its processes, and its evolution, through the use of simulations with computer models, robotics, and biochemistry. The discipline was named by Christopher Langton, an American theoretical biologist, in 1986. In 1987 Langton organized the first conference on the field, in Los Alamos, New Mexico. There are three main kinds of ALife named for their approaches: soft, from software; hard, from hardware; and wet, from biochemistry. Artificial Life researchers study traditional biology by trying to recreate aspects of biological phenomena. The modelling philosophy of Artificial Life strongly differs from traditional modelling by studying not only "life-as-we-know-it" but also "life-as-it-might-be".

A traditional model of a biological system will focus on capturing its most important parameters. In contrast, an Alife modelling approach will generally seek to decipher the most simple and general principles underlying life and implement them in a simulation. The simulation then offers the possibility to analyse new and different lifelike systems. [Vladimir Georgievich Red'ko](#) proposed to generalize this distinction to the modelling of any process, leading to the more general distinction of "processes-as-we-know-them" and "processes-as-they-could-be". At present, the commonly accepted definition of life does not consider any current Alife simulations or software to be Alive, and they do not constitute part of the evolutionary process of any ecosystem. However, different opinions about Artificial Life's potential have arisen: The strong Alife (cf. Strong AI) position states that "life is a process which can be abstracted away from any particular medium" (John von Neumann). Notably, Tom Ray declared that his program Tierra is not simulating life in a computer but synthesizing it. The weak Alife position denies the possibility of generating a "living process" outside of a chemical solution. Its researchers try instead to simulate life processes to understand the underlying mechanics of biological phenomena."

32.1.2

Can you Remember? What is meant by 'Artificial Life' and what could it add the current AI systems?

32.1.3

Question: Vladimir Red'ko poses the preference not to try and emulate the virtues of natural evolution like the human brains, but better try to study "processes-as-they-could-be". In other words, to excavate the mechanisms underlying natural evolution ('wet' intelligence) so that we are less obsessed (and thus 'hampered') by current silicon and software limitations. What alternative approaches come to your mind for actively undertaking the suggestion posed by Vladimir Red'ko? In case you have no idea, please take into account notions from other literature like: 'self-organisation', 'quantum computing' and 'wisdom of the crowd'.

AI and Creativity

Chapter **33**

33.1 Artificial Intelligence and Creativity

33.1.1

Artificial Intelligence and Creativity

TEDx: How AI shapes a new form of creativity.



“Dr. Johannes Stelzer has focused on the transfer of AI advanced technology into the application. After studying physics, he spent over a decade researching at various Max Planck Institutes at the intersection of neuroscience and artificial intelligence. He sees particularly high potential in the new phenomenon of creative artificial intelligence. Johannes Stelzer is a founding member of the artist collective Lunar Ring and makes creative AI tangible in the context of exhibitions and events.”

According to Wikipedia: “ .. [Jürgen Schmidhuber](#)'s formal theory of creativity postulates that creativity, curiosity, and interestingness are by-products of a simple computational principle for measuring and optimizing learning progress. Consider an agent able to manipulate its environment and thus its own sensory inputs. The agent can use a black box optimization method such as reinforcement learning to learn (through informed trial and error) sequences of actions that maximize the expected sum of its future reward signals. There are extrinsic reward signals for achieving externally given goals, such as finding food when hungry. This non-standard term

motivates purely creative behaviour of the agent even when there are no external goals. A wow-effect is formally defined as follows. As the agent is creating, predicting and encoding the continually growing history of actions and sensory inputs, it keeps improving the predictor or encoder, which can be implemented as an artificial neural network or some other machine learning device that can exploit regularities in the data to improve its performance over time..”

33.1.2

Can you Remember? What means Jürgen Schmidhuber as he postulates the ‘formal theory of creativeness’? What would be the equivalent for ‘creative AI’?

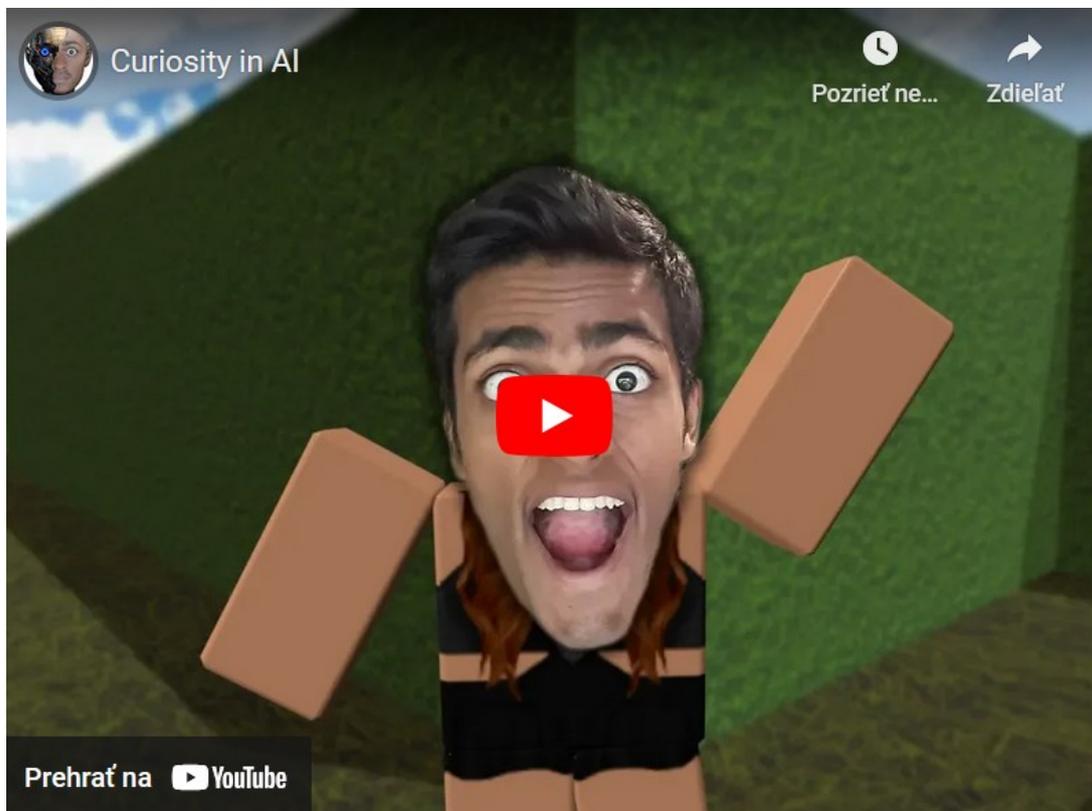
33.1.3

Question: If you explore Schmidhuber’s theory on rudimentary artificial science, what contribution to future AI would you see in creative art? In other words: What contribution to current AI was given by Science Fiction in the 19th and mid 20th century (visionary writers like Jules Verne, H. G. Wells and Edgar Allan Poe)?

33.2 Curiosity in Artificial Intelligence

33.2.1

[Curiosity in Artificial Intelligence](#)

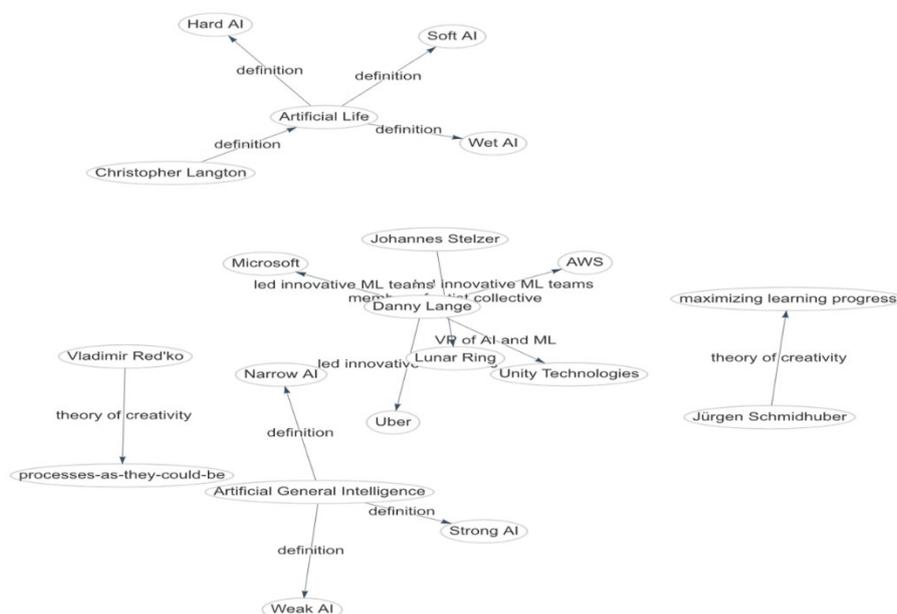


“ Reinforcement learning generally uses a carrot-and-stick approach. Good actions are rewarded, and bad actions are punished. But what are the drawbacks of this simple approach? How can we use curiosity to overcome it? Let's find out in this video!”

According to [Wikipedia](#): “ .. Curiosity can be seen as an innate quality of many different species. It is common to human beings at all ages from infancy through adulthood and is easy to observe in many other animal species; these include apes, cats, and rodents. Early definitions cite curiosity as a motivated desire for information. This motivational desire has been said to stem from a passion or an appetite for knowledge, information, and understanding. These traditional ideas of curiosity have recently expanded to look at the difference between perceptual curiosity as the innate exploratory behaviour that is present in all animals and epistemic curiosity as the desire for knowledge that is specifically attributed to humans.

Daniel Berlyne recognized three major variables playing a role in evoking curiosity: psychophysical variables, ecological variables, and collative variables. Psychophysical variables correspond to physical intensity, while ecological variables to motivational significance and task relevance. Collative variables are called “collative” because they involve a comparison between different stimuli or features, which may be actually perceived or which may be recalled from memory. Berlyne mentioned four collative variables: novelty, complexity, uncertainty, and conflict. At the same time, he suggested that all collative variables probably involve conflict. Additionally, he considered three variables supplementary to novelty: change, surprisingness, and incongruity. Finally, curiosity may not only be aroused by the perception of some stimulus associated with the aforementioned variables (“specific exploration”), but also by a lack of stimulation, out of “boredom” (“diversive exploration”).”

Key Players in the Triangle Microsoft, Uber and Learning Creativity



33.2.2

Can you Remember? Which three major variables play a crucial role in the arousal of 'curiosity' due to Berlyne?

33.2.3

Question: Berlyne's postulate of 'collative variables' brought him to the construct of 'conflict'. Does it explain that travelling raises more new ideas than sitting behind your desk? Talking about curiosity towards AI; What aspects of future AI attracts your imagination most?

33.3 Artificial Intelligence in Video Games

33.3.1

Artificial Intelligence in Video Games



"AI in video games is very different than the AI in self-driving cars, smart home devices, and natural language algorithms. That's because game developers have been sticking to the basics, but all that is starting to change. Researchers today are incorporating more cutting-edge AI advancements into games. In the future, these

tools could result in games designing themselves with characters that can learn and adapt to the player.”

According to [Wikipedia](#): “ .. In video games, artificial intelligence (AI) is used to generate responsive, adaptive or intelligent behaviors primarily in non-player characters (NPCs) similar to human-like intelligence. Artificial intelligence has been an integral part of video games since their inception in the 1950s. AI in video games is a distinct subfield and differs from academic AI. It serves to improve the game-player experience rather than machine learning or decision making. During the golden age of arcade video games, the idea of AI opponents was largely popularized in the form of graduated difficulty levels, distinct movement patterns, and in-game events depend on the player's input. Modern games often implement existing techniques such as pathfinding and decision trees to guide the actions of NPCs. AI is often used in mechanisms which are not immediately visible to the user, such as data mining and procedural-content generation. “

33.3.2

Can you Remember? What is the major role of AI in designing Video Games?

33.3.3

Question: Games encourage humans to think about the future as they pose fiction as temporary reality first of all. The term ‘Virtual Reality’ as ‘fictitious world’ slowly becomes accepted as the default 3D space that overcomes physical alienation as we experienced during video conferencing during the COVID-19 era. What means virtual reality for you; is it just a corollary of gaming and thus a way of escaping from full reality; or is it already an innate need for current teenagers?

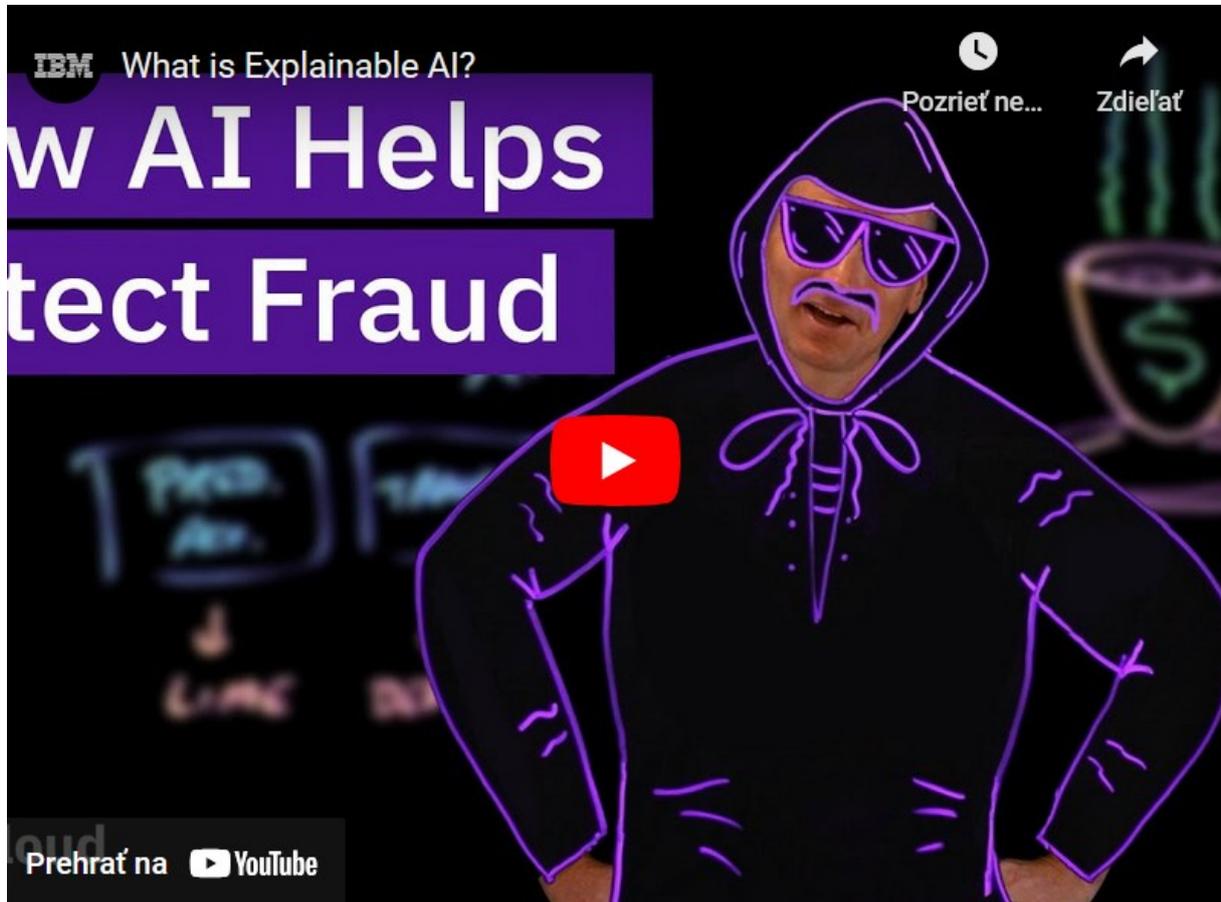
Explainable AI

Chapter **34**

34.1 Explainable Artificial Intelligence

34.1.1

[Explainable Artificial Intelligence](#)



" Explainable artificial intelligence (XAI) is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms. Explainable AI is used to describe an AI model, its expected impact and potential biases. Sound complicated? Master Inventor Martin Keen gives you a simple (and fun) explanation on how explainable AI works."

According to [Wikipedia](#): " .. Explainable AI (XAI), or Interpretable AI, or Explainable Machine Learning (XML), is artificial intelligence (AI) in which humans can understand the decisions or predictions made by the AI. It contrasts with the "black box" concept in machine learning where even its designers cannot explain why an AI arrived at a specific decision. By refining the mental models of users of AI-powered systems and dismantling their misconceptions, XAI promises to help users perform more effectively. XAI may be an implementation of the social right to explanation. XAI is relevant even if there is no legal right or regulatory requirement. For example, XAI can improve the user experience of a product or service by helping end users trust that the AI is making good decisions. This way the aim of XAI is to explain what has been done, what is done right now, what will be done next and unveil the

information the actions are based on. These characteristics make it possible (i) to confirm existing knowledge (ii) to challenge existing knowledge and (iii) to generate new assumptions.

The algorithms used in AI can be differentiated into white-box and black-box machine learning (ML) algorithms. White-box models are ML models that provide results that are understandable for experts in the domain. Black-box models, on the other hand, are extremely hard to explain and can hardly be understood even by domain experts. XAI algorithms are considered to follow the three principles of transparency, interpretability and explainability. Transparency is given “if the processes that extract model parameters from training data and generate labels from testing data can be described and motivated by the approach designer”. Interpretability describes the possibility of comprehending the ML model and presenting the underlying basis for decision-making in a way that is understandable to humans.

Explainability is a concept that is recognized as important, but a joint definition is not yet available. It is suggested that explainability in ML can be considered as “the collection of features of the interpretable domain, that have contributed for a given example to produce a decision (e.g., classification or regression)”. If algorithms meet these requirements, they provide a basis for justifying decisions, tracking and thereby verifying them, improving the algorithms, and exploring new facts. Sometimes it is also possible to achieve a result with high accuracy with a white-box ML algorithm that is interpretable in itself. This is especially important in domains like medicine, defense, finance and law, where it is crucial to understand the decisions and build up trust in the algorithms. Many researchers argue that, at least for supervised machine learning, the way forward is symbolic regression, where the algorithm searches the space of mathematical expressions to find the model that best fits a given dataset. AI systems optimize behaviour to satisfy a mathematically specified goal system chosen by the system designers, such as the command “maximize accuracy of assessing how positive film reviews are in the test dataset”. The AI may learn useful general rules from the test set, such as “reviews containing the word ‘horrible’ are likely to be negative”. However, it may also learn inappropriate rules, such as “reviews containing ‘Daniel Day-Lewis’ are usually positive”; such rules may be undesirable if they are deemed likely to fail to generalize outside the train set, or if people consider the rule to be “cheating” or “unfair”. A human can audit rules in an XAI to get an idea how likely the system is to generalize to future real-world data outside the test-set.”

34.1.2

Can you Remember? What is the role of ‘Explainable AI’ (XAI), and how has this challenge been approached so far?

34.1.3

Question: Supervised machine learning allows the expert to stepwise regulate the amount of variables and dimensions to be taken into account in order to still ‘understand’ a certain AI conclusion. If you think of your own (human) personal learning in the past, you might need to accept that parts of your academic learning

was not always 'transparent': you were not always able to understand the building blocks that you use for 'seeing the larger picture'. For instance during COVID-19 it was not easy to understand how viruses propagated and how immunity developed. Can you go back to your own understanding of virology and confess what pervasive misconceptions you needed to give up finally?

34.2 Artificial Intelligence Takeover

34.2.1

Artificial Intelligence Takeover

According to [Wikipedia](#): “ .. An AI takeover is a hypothetical scenario in which an artificial intelligence (AI) becomes the dominant form of intelligence on Earth, as computer programs or robots effectively take the control of the planet away from the human species. Possible scenarios include replacement of the entire human workforce, takeover by a super intelligent AI, and the popular notion of a robot uprising. Some public figures, such as Stephen Hawking and Elon Musk, have advocated research into precautionary measures to ensure future super intelligent machines remain under human control. The traditional consensus among economists has been that technological progress does not cause long-term unemployment.

However, recent innovation in the fields of robotics and artificial intelligence has raised worries that human labour will become obsolete, leaving people in various sectors without jobs to earn a living, leading to an economic crisis. Many small and medium size businesses may also be driven out of business if they will not be able to afford or license the latest robotic and AI technology and may need to focus on areas or services that cannot easily be replaced for continued viability in the face of such technology. Technologies that may displace workers. AI technologies have been widely adopted in recent years, and this trend will only continue to gain popularity given the digital transformation efforts from companies across the world.

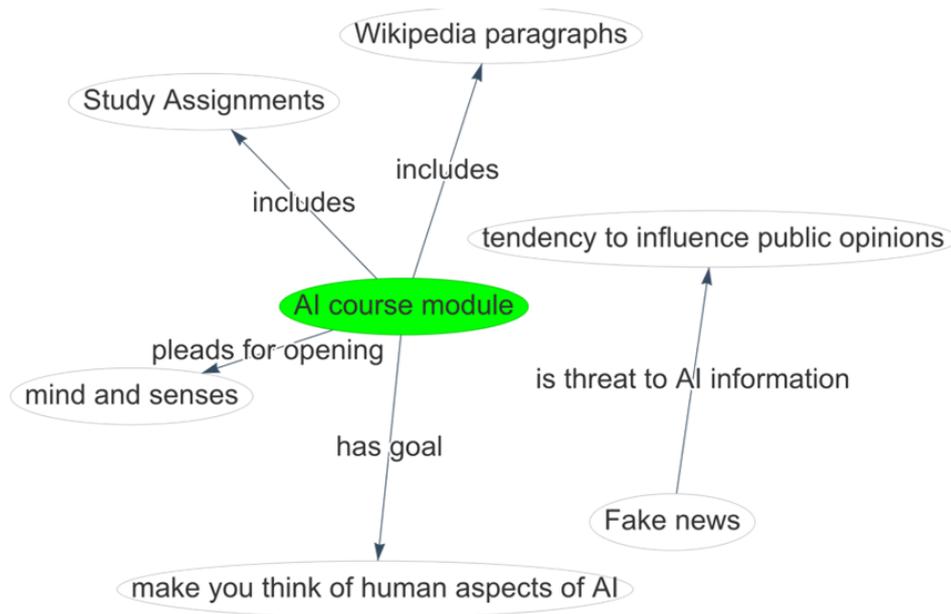
While these technologies have replaced many traditional workers, they also create new opportunities. Industries that are most susceptible to experience AI takeover include transportation, retail, and military. AI military technologies, for example, allow soldiers to work remotely without any risk of injury.

1. **Author Dave Bond** argues that as AI technologies continue to develop and expand, the relationship between humans and robots will change; they will become closely integrated in several aspects of life. Overall, it is safe to assume that AI will displace some workers while creating opportunities for new jobs in other sectors, especially in fields where tasks are repeatable.
2. **Computer-integrated manufacturing.** Computer-integrated manufacturing is the manufacturing approach of using computers to control the entire production process. This integration allows individual processes to exchange information with each other and initiate actions. Although

manufacturing can be faster and less error-prone by the integration of computers, the main advantage is the ability to create automated manufacturing processes. Computer-integrated manufacturing is used in automotive, aviation, space, and ship building industries.

3. **White-collar machines.** The 21st century has seen a variety of skilled tasks partially taken over by machines, including translation, legal research and even low-level journalism. Care work, entertainment, and other tasks requiring empathy, previously thought safe from automation, have also begun to be performed by robots.
4. **Autonomous cars.** An autonomous car is a vehicle that is capable of sensing its environment and navigating without human input. Many such vehicles are being developed, but as of May 2017 automated cars permitted on public roads are not yet fully autonomous. They all require a human driver at the wheel who is ready at a moment's notice to take control of the vehicle. Among the main obstacles to widespread adoption of autonomous vehicles, are concerns about the resulting loss of driving-related jobs in the road transport industry. On March 18, 2018, the first human was killed by an autonomous vehicle in Tempe, Arizona by an Uber self-driving car.
5. **Eradication.** Scientists such as Stephen Hawking are confident that superhuman artificial intelligence is physically possible, stating "there is no physical law precluding particles from being organized in ways that perform even more advanced computations than the arrangements of particles in human brains". Scholars like Nick Bostrom debate how far off superhuman intelligence is, and whether it would actually pose a risk to mankind. According to Bostrom, a super intelligent machine would not necessarily be motivated by the same emotional desire to collect power that often drives human beings but might rather treat power as a means toward attaining its ultimate goals; taking over the world would both increase its access to resources and help to prevent other agents from stopping the machine's plans. As an oversimplified example, a paperclip maximiser designed solely to create as many paperclips as possible would want to take over the world so that it can use all of the world's resources to create as many paperclips as possible, and, additionally, prevent humans from shutting it down or using those resources on things other than paperclips."

Fake News becoming a Threat for A.I. as well?



 34.2.2

Can you Remember? The Eradication scenario is a dramatic version of the AI take-over. Which one of the other four is the most influential one? Please argue why.

 34.2.3

Question: Stephen Hawking, based upon a cosmic humility, claimed that there is no physical law precluding particles from being organized in ways that perform even more advanced computations than the arrangements of particles in human brains. What is your opinion on our progress to 'connect human minds' via 'electrodes' and 'future social media'? What options are there for future AI to play a stimulating role in this?

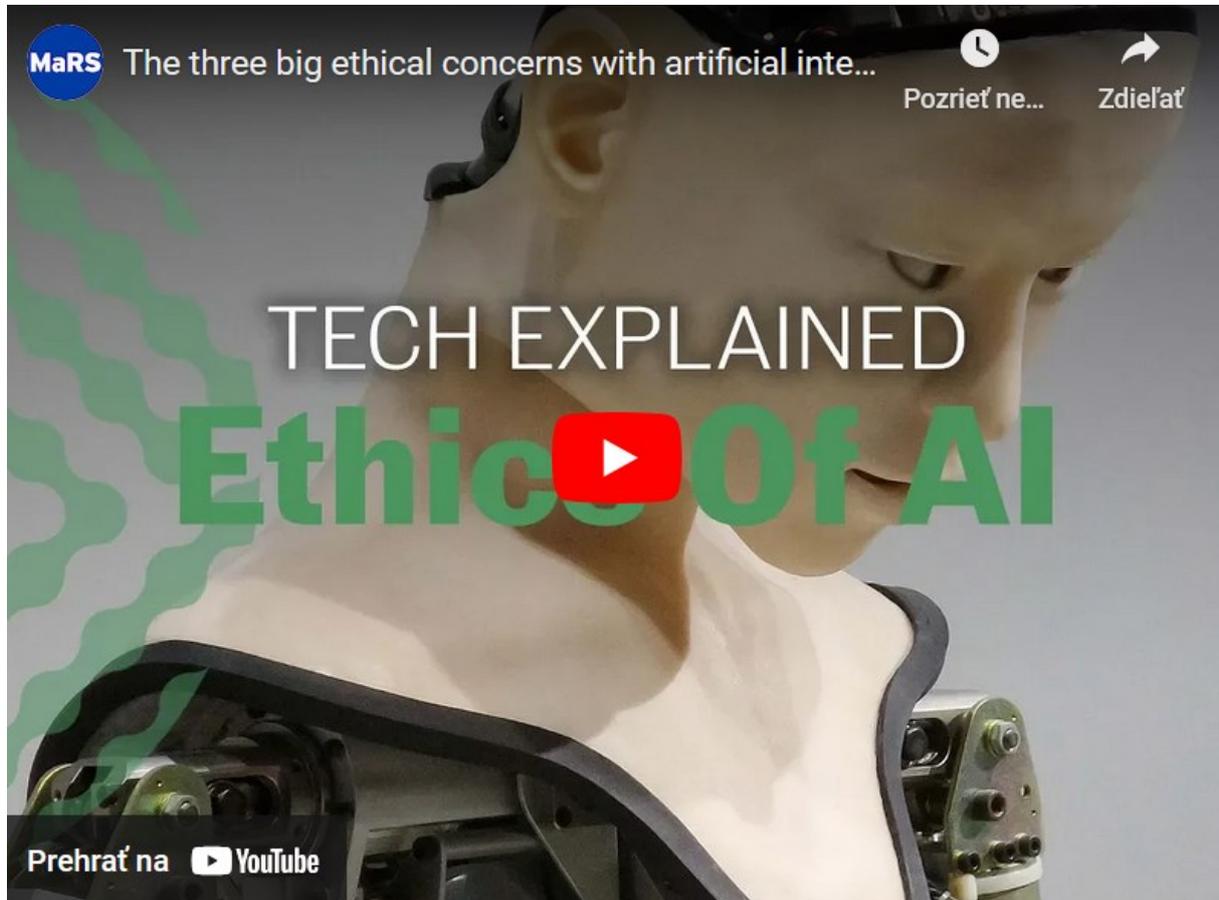
Ethical Aspects of AI

Chapter **35**

35.1 Ethical Aspects of Artificial Intelligence

35.1.1

[Ethical Aspects of Artificial Intelligence](#)



“ Frank Rudzicz is an artificial intelligence researcher at the University of Toronto and at the Vector Institute. The Vector Institute is an independent, not-for-profit corporation dedicated to research in the field AI, with special focus in machine and deep learning.”

According to Wikipedia: “ .. Artificial consciousness (AC), also known as machine consciousness (MC) or synthetic consciousness (Gamez 2008; Reggia 2013), is a field related to artificial intelligence and cognitive robotics. The aim of the theory of artificial consciousness is to "Define that which would have to be synthesized were consciousness to be found in an engineered artifact" (Aleksander 1995). Neuroscience hypothesizes that consciousness is generated by the interoperation of various parts of the brain, called the Neural Correlates of Consciousness or NCC, though there are challenges to that perspective. Proponents of AC believe it is possible to construct systems (e.g., computer systems) that can emulate this NCC interoperation. Artificial consciousness concepts are also pondered in the philosophy of artificial intelligence through questions about mind, consciousness, and mental states. Type-identity theorists and other sceptics hold the view that

consciousness can only be realized in particular physical systems because consciousness has properties that necessarily depend on physical constitution (Block 1978; Bickle 2003).

In his article "Artificial Consciousness: Utopia or Real Possibility," Giorgio Buttazzo says that a common objection to artificial consciousness is that "Working in a fully automated mode, they cannot exhibit creativity, un-reprogrammation (which means can no longer be reprogrammed, from rethinking), emotions, or free will. A computer, like a washing machine, is a slave operated by its components." For other theorists (e.g., functionalists), who define mental states in terms of causal roles, any system that can instantiate the same pattern of causal roles, regardless of physical constitution, will instantiate the same mental states, including consciousness (Putnam 1967).

Ethics

If it were suspected that a particular machine was conscious, its rights would be an ethical issue that would need to be assessed (e.g. what rights it would have under law). For example, a conscious computer that was owned and used as a tool or central computer of a building of larger machine is a particular ambiguity. Should laws be made for such a case? Consciousness would also require a legal definition in this particular case. Because artificial consciousness is still largely a theoretical subject, such ethics have not been discussed or developed to a great extent, though it has often been a theme in fiction. In 2021, the German philosopher Thomas Metzinger has demanded a global moratorium on synthetic phenomenology until 2050, on ethical grounds. The rules for the 2003 Loebner Prize competition explicitly addressed the question of robot rights: If, in any given year, a publicly available open-source Entry entered by the University of Surrey or the Cambridge Center wins the Silver Medal or the Gold Medal, then the Medal and the Cash Award will be awarded to the body responsible for the development of that Entry. If no such body can be identified, or if there is disagreement among two or more claimants, the Medal and the Cash Award will be held in trust until such time as the Entry may legally possess, either in the United States of America or in the venue of the contest, the Cash Award and Gold Medal in its own right..."

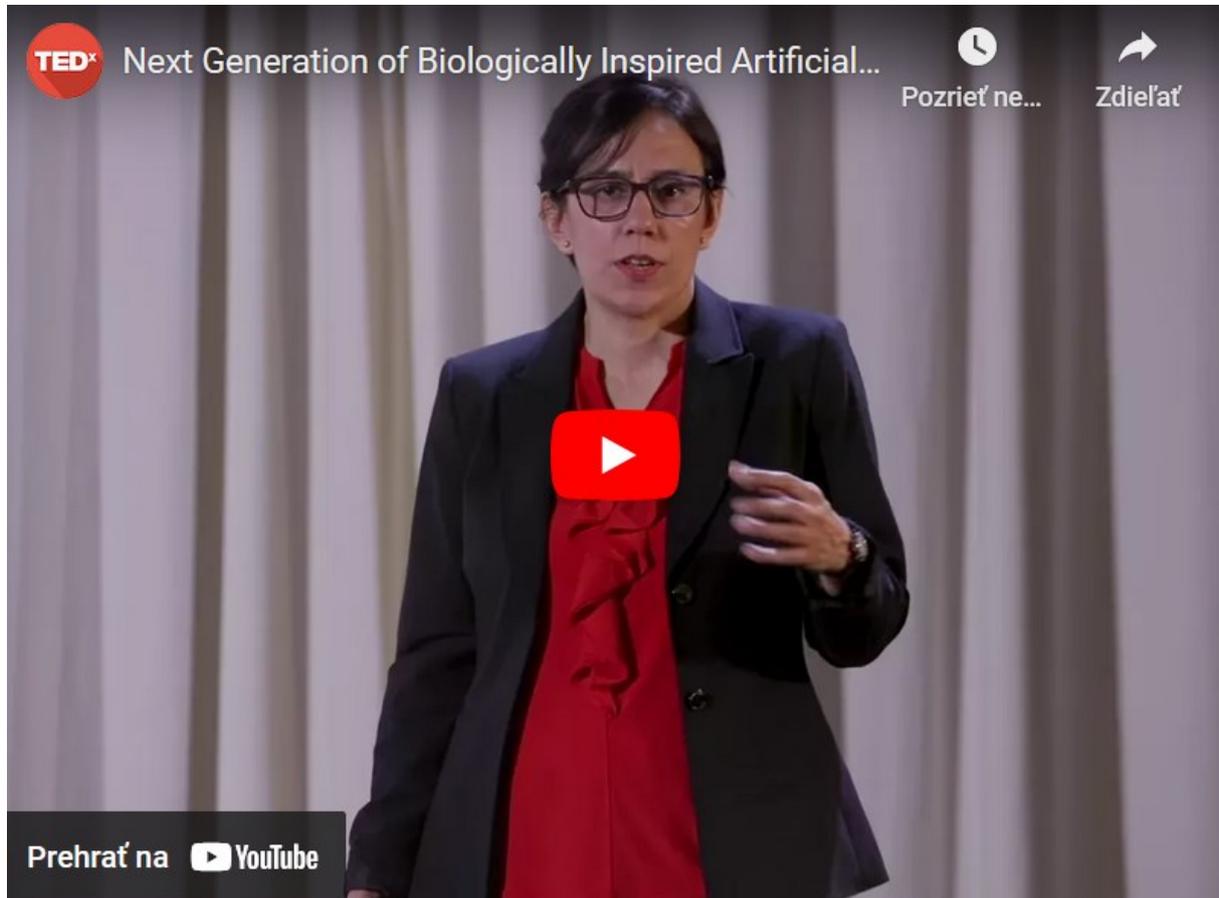
35.1.2

Can you Remember? Neural Correlates of Consciousness or NCC has been postulated to explain how human awareness evolves. However NCC heavily relies on a wide spectrum of physiological bodily processes. Would it be a good idea to equip AI consciousness with a complete emulation of the human body: respiration, heart-beat, hormones etc.? Is AI consciousness an obligatory ingredient before AI ethics can become achieved?

35.2 Bio-Inspired Artificial Intelligence

35.2.1

Bio-Inspired Artificial Intelligence



TEDx: Next Generation of Biologically Inspired Artificial Intelligence.

“ ...Simulation of mechanisms of coding, data processing in biological systems at the molecular level provides promising tools for next generation cognitive systems with the capability of real-time response to the environmental signals. In large research programs, there is biological coding to define algorithms for solving NP hard problems. This TED Talk presents the results and provide insights and applications for molecular sensors and implications for next generation of artificial intelligence. Dr. Tara Karimi is a multi-disciplinary scientist who has devoted her academic life to learning biological systems and applying the natural principles to the outside world. Dr. Karimi holds two PhDs in veterinary science and biochemistry and has completed several post-doc projects in tissue engineering, genetic engineering, molecular and developmental biology, stem cell research, and regenerative medicine.”

According to [Wikipedia](#): “ .. The ideas behind biological computing trace back to 1936 and the first description of an abstract computer, which is now known as a Turing

machine. Turing firstly described the abstract construct using a biological specimen. Turing imagined a mathematician that has three important attributes. He always has a pencil with an eraser, an unlimited number of papers and a working set of eyes. The eyes allow the mathematician to see and perceive any symbols written on the paper while the pencil allows him to write and erase any symbols that he wants. Lastly, the unlimited paper allows him to store anything he wants memory. Using these ideas he was able to describe an abstraction of the modern digital computer. However, Turing mentioned that anything that can perform these functions can be considered such a machine and he even said that even electricity should not be required to describe digital computation and machine thinking in general...”

35.2.2

Can you Remember? What is the most elementary imagination of the computer based upon the [Von Neumann architecture](#)?

35.2.3

Neural Networks

First described in 1943 by Warren McCulloch and Walter Pitts, neural networks are a prevalent example of biological systems inspiring the creation of computer algorithms. They first mathematically described that a system of simplistic neurons was able to produce simple logical operations such as logical conjunction, disjunction, and negation. They further showed that a system of neural networks can be used to carry out any calculation that requires finite memory. Around 1970 the research around neural networks slowed down and many consider the 1969 book “Perceptrons: An Introduction to Computational Geometry” by Marvin Minsky and Seymour Papert as its main cause.

Their book showed that neural network models were able only to model systems that are based on Boolean functions that are true only after a certain threshold value. Such functions are also known as threshold functions. The book also showed that a large number of systems cannot be symbolically represented as such means that a large amount of systems cannot be modelled by neural networks. Another book by James Rumelhart and David McClelland in 1986 brought neural networks back to the spotlight by demonstrating the linear back-propagation algorithm, something that allowed the development of multi-layered neural networks.

35.2.4

Can you Remember? The 1943 choice to limit Neuronal computing to binary logic has been quite decisive for AI control structures since then. [Quantum Computing](#) (QC) offers a new dimension. Can you imagine what it QC may imply for AI finally?

35.2.5

Ant Colonies

Douglas Hofstadter in 1979 described an idea of a biological system capable of performing intelligent calculations even though the individuals comprising the system might not be intelligent. More specifically, he gave the example of an ant colony that can carry out intelligent tasks together but each individual ant cannot exhibit something called "emergent behaviour." Azimi et al. in 2009 showed that what they described as the "ant colony" algorithm, is based upon a clustering algorithm that is able to produce highly competitive final clusters comparable to other traditional algorithms. Lastly Hölder and Wilson in 2009 concluded, using historical data, that ants have evolved to function as a single "superorganism" colony. A very important result since it suggested that group selection evolutionary algorithms coupled together with algorithms similar to the "ant colony" can be potentially used to develop more powerful algorithms.

35.2.6

Can you Remember? Can you recollect why the metaphor of ant colony triggers the imagination of AI by parallel processing?

35.2.7

Question: Turing defined the needed tools for mathematics: paper, pencil, a working set of eyes and eraser. One could see a one-to-one isomorphy to the Neumann architecture where the processing unit reads a linear sequence of data and operations that could even make the reading on addresses forward and backward. The later programming paradigms attempted to stay at the level of 'rules' and 'facts' so that procedural prescriptions became minimal. If you have some experience with declarative languages like LISP or PROLOG, please provide an example that shows that procedural awareness is still needed.

35.3 Existential Risk from Artificial General Intelligence

35.3.1

Existential Risk from Artificial General Intelligence

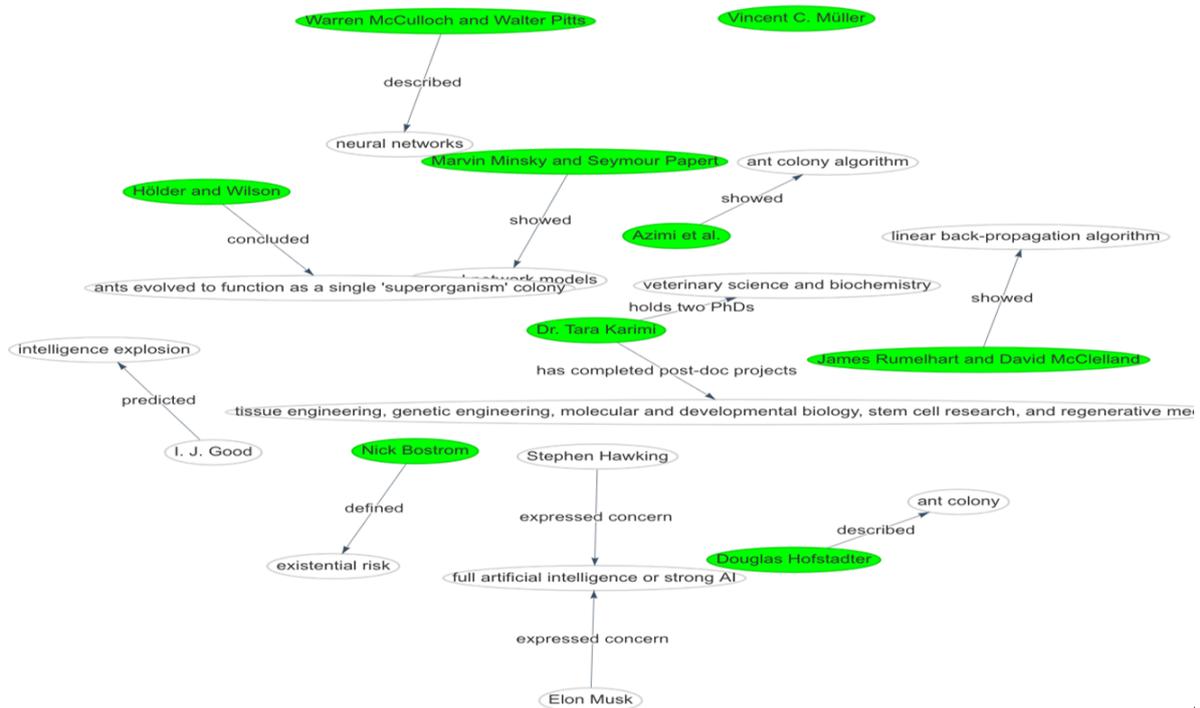


“ Artificial superintelligence or a superintelligence in general is a hypothetical agent that possesses intelligence far surpassing that of the brightest and most gifted human minds. According to the most popular version of the singularity hypothesis, called intelligence explosion, an upgradable intelligent agent will eventually enter a "runaway reaction" of self-improvement cycles, each new and more intelligent generation appearing more and more rapidly, causing an "explosion" in intelligence, and resulting in a powerful superintelligence that qualitatively far surpasses all human intelligence. I. J. Good's "intelligence explosion" model predicts that a future superintelligence will trigger a singularity. Four polls of AI researchers, conducted in 2012 and 2013 by Nick Bostrom and Vincent C. Müller, suggested a median probability estimate of 50% that artificial general intelligence or AGI would be developed by 2040 to 2050. Public figures such as Stephen Hawking and Elon Musk have expressed concern that full artificial intelligence or strong AI could result in human extinction. But the consequences of the singularity and its potential benefit or harm to the human race have been intensely debated. Philosopher Nick Bostrom

defines an existential risk as one in which an extinction-level event is not only possible but likely. He argues that advanced artificial intelligence is "likely" to be an existential risk. In addition, time frame is also a factor, as a superintelligence might decide to act quickly before humans would have a chance to react with any countervailing action. A superintelligence might decide to pre-emptively eliminate all of humanity for reasons that may be incomprehensible to us. There is also a possibility where a superintelligence might seek to colonize the universe. A superintelligence might do this in order to maximize the amount of computation it could do or to obtain raw materials for manufacturing new supercomputers. While it may seem alarmist to worry about these scenarios in a current world where only narrow AI exists, we do not know how long it takes or if it's even possible to develop safe artificial super-intelligence that shares our goals. Thus, we better start planning today about the advent of ASI... While we still can! ..."

According to [Wikipedia](#): " .. It is argued that the human species currently dominates other species because the human brain has some distinctive capabilities that other animals lack. If AI surpasses humanity in general intelligence and becomes "super intelligent", then it could become difficult or impossible for humans to control. Just as the fate of the mountain gorilla depends on human goodwill, so might the fate of humanity depend on the actions of a future machine superintelligence. The chance of this type of scenario is widely debated, and hinges in part on differing scenarios for future progress in computer science. Once the exclusive domain of science fiction, concerns about superintelligence started to become mainstream in the 2010s, and were popularized by public figures such as Stephen Hawking, Bill Gates, and Elon Musk. One source of concern is that controlling a super intelligent machine, or instilling it with human-compatible values, may be a harder problem than naïvely supposed. Many researchers believe that a superintelligence would naturally resist attempts to shut it off or change its goals—a principle called instrumental convergence—and that pre-programming a superintelligence with a full set of human values will prove to be an extremely difficult technical task. In contrast, sceptics such as computer scientist Yann LeCun argue that super intelligent machines will have no desire for self-preservation. A second source of concern is that a sudden and unexpected "intelligence explosion" might take an unprepared human race by surprise. To illustrate, if the first generation of a computer program able to broadly match the effectiveness of an AI researcher, is able to rewrite its algorithms and double its speed or capabilities in six months, then the second-generation program is expected to take three calendar months to perform a similar chunk of work. In this scenario the time for each generation continues to shrink, and the system undergoes an unprecedentedly large number of generations of improvement in a short time interval, jumping from subhuman performance in many areas to superhuman performance in all relevant areas. Empirically, examples like AlphaZero in the domain of Go show that AI systems can sometimes progress from narrow human-level ability to narrow superhuman ability extremely rapidly..."

A.I. Heroes, standing on the Shoulders of early Pioneers:



35.3.2

Can you Remember? What could AI offer to protect nature? Less travelling? Less? Less spoiled energy? Less wars? Please formulate your thoughts based upon the prior video and Wikipedia fragment.

35.3.3

Question: Algorithmic programming allowed to iterate quit fast along computations, even entire matrix operations could be anticipated within one keystroke by the programmer, for instance in the IBM language APL. Later efforts evolved to base programming upon functional-, object-oriented- and aspect-oriented languages. If you monitor the trend in programming styles, do you see an indication that we get closer and closer to the needs in AI applications?

35.4 Embodied Cognition

35.4.1

Embodied Cognition



Some review outcomes: "Wow Karl, you just explained material worth 10 books in such a compelling, comprehensive, and cogent way. Thanks a ton. I was looking for something like this as a researcher in problem solving and situated cognition." And: " What a great illustration navigating through multiple-complex ideas with incredible agility and eloquence. Please we want more of him."

Embodied Cognition rests upon the paradigm that essentially any element in thinking and mental imagination rests upon the physical substrata in the human actor. And as such it is mind-boggling how AI can work, based upon silicon and programmed algorithms. According to [Wikipedia](#): " .. Embodied cognition is the theory that many features of cognition, whether human or otherwise, are shaped by aspects of an organism's entire body. Sensory and motor systems are seen as fundamentally integrated with cognitive processing. The cognitive features include high-level mental constructs (such as concepts and categories) and performance on various cognitive tasks (such as reasoning or judgment). The bodily aspects involve the motor system, the perceptual system, the bodily interactions with the environment (situatedness), and the assumptions about the world built into the organism's

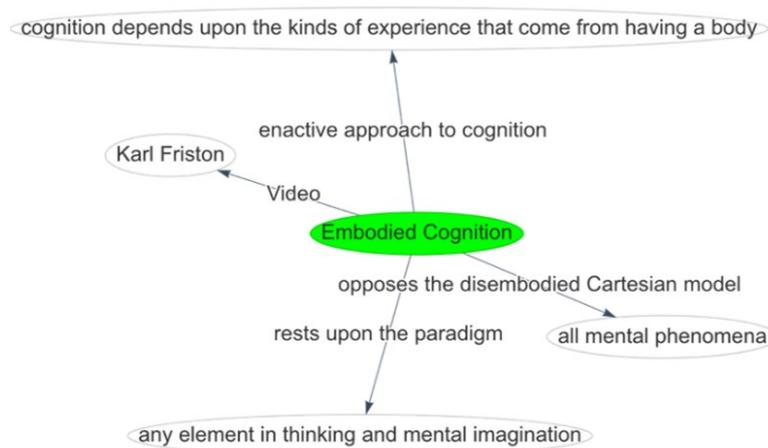
functional structure. The embodied mind thesis challenges other theories, such as cognitivism, computationalism, and Cartesian dualism. It is closely related to the extended mind thesis, situated cognition, and enactivism. The modern version depends on insights drawn from up-to-date research in psychology, linguistics, cognitive science, dynamical systems, artificial intelligence, robotics, animal cognition, plant cognition, and neurobiology. Proponents of the embodied cognition thesis emphasize the active and significant role the body plays in the shaping of cognition and in the understanding of an agent's mind and cognitive capacities..."

 35.4.2

Question: The mind-body dualism as the basis for mentalistic AI easily leads to ignore situational and bodily factors. Will AI perform better if these factors are included in the computational model?

 35.4.3

Embodied Cognition as Source for Wet A.I.:



In philosophy, embodied cognition holds that an agent's cognition, rather than being the product of mere (innate) abstract representations of the world, is strongly influenced by aspects of an agent's body beyond the brain itself. An embodied model of cognition opposes the disembodied Cartesian model, according to which all mental phenomena are non-physical and, therefore, not influenced by the body. With this opposition the embodiment thesis intends to reintroduce an agent's bodily experiences into any account of cognition. It is a rather broad thesis and encompasses both weak and strong variants of embodiment. In an attempt to reconcile cognitive science with human experience, the enactive approach to cognition defines "embodiment" as follows: By using the term embodied we mean to highlight two points: first that cognition depends upon the kinds of experience that come from having a body with various sensorimotor capacities, and second, that these individual sensorimotor capacities are themselves embedded in a more encompassing biological, psychological and cultural context (The Embodied Mind: Cognitive Science and Human Experience by Francisco J. Varela, Evan Thompson, and Eleanor Rosch, pages 172–173.)

35.4.4

Question: Environmental-, diverse-, emancipatory-, democratic- and transparent are all buzz words in Western Society the last two decades. What do these trends imply for next generations AI?

35.4.5

This double sense attributed to the embodiment thesis emphasizes the many aspects of cognition that researchers in different fields - such as philosophy, cognitive science, artificial intelligence, psychology, and neuroscience—are involved with. This general characterization of embodiment faces some difficulties: a consequence of this emphasis on the body, experience, culture, context, and the cognitive mechanisms of an agent in the world is, that often-distinct views and approaches to embodied cognition overlap. The theses of extended cognition and situated cognition, for example, are usually intertwined and not always carefully separated. And since each of the aspects of the embodiment thesis is endorsed to different degrees, embodied cognition should be better seen "as a research program rather than a well-defined unified theory". Some authors explain the embodiment thesis by arguing that cognition depends on an agent's body and its interactions with a determined environment. From this perspective, cognition in real biological systems is not an end in itself, it is constrained by the system's goals and capacities. It would be a mistake, however, to suppose that cognition consists simply of building maximally accurate representations of input information. The gaining of knowledge is a steppingstone to achieving the more immediate goal of guiding behaviour in response to the system's changing surroundings (Marcin Miłkowski, Explaining the Computational Mind, p. 4.). The embodied cognitive model of the mind under which body, world, perception, and action are dynamically related with each other.

35.4.6

Can you Remember? “ .. It would be a mistake, however, to suppose that cognition consists simply of building maximally accurate representations of input information” Does this mean that anthropomorph robots are no longer needed?

35.4.7

Another approach to understanding embodied cognition comes from a narrower characterization of the embodiment thesis. The following narrower view of embodiment avoids any compromises to external sources other than the body and allows differentiating between embodied cognition, extended cognition, and situated cognition. Thus, the embodiment thesis can be specified as follows: Many features of cognition are embodied in that they are deeply dependent upon characteristics of the physical body of an agent, such that the agent's beyond-the-brain body plays a significant causal role, or a physically constitutive role, in that agent's cognitive processing (RA Wilson and L Foglia, Embodied Cognition in the Stanford Encyclopaedia of Philosophy.). This thesis points out the core idea that an agent's

body plays a significant role in shaping different features of cognition, such as perception, attention, memory, reasoning - among others. Likewise, these features of cognition depend on the kind of body an agent has. In 2002, a collection of positive characterizations summarizing what the embodiment thesis entails for cognition were offered. Margaret Wilson argues that the general outlook of embodied cognition "displays an interesting co-variation of multiple observations and houses a number of different claims:

1. cognition is situated;
2. cognition is time-pressured;
3. we off-load cognitive work onto the environment;
4. the environment is part of the cognitive system;
5. cognition is for action;
6. offline cognition is bodily-based".

According to Wilson, the first three and the fifth claim appear to be at least partially true, while the fourth claim is deeply problematic in that all things that have an impact on the elements of a system are not necessarily considered part of the system. The sixth claim has received the least attention in the literature on embodied cognition, yet it might be the most significant of the six claims as it shows how certain human cognitive capabilities, that previously were thought to be highly abstract, now appear to be leaning towards an embodied approach for their explanation. Wilson also describes at least five main (abstract) categories that combine both sensory and motor skills (or sensorimotor functions). The first three are working memory, episodic memory, and implicit memory; the fourth is mental imagery, and finally, the fifth concerns reasoning and problem solving.

35.4.8

Question: When we accept bodily sensation and emotions as key factor in human cognition, we arrive closely to the role of 'intuition' in problem solving tasks. For instance, 'estimating' the size of large areas may easily become blurred as its magnitude stretch far beyond what we can normally see, hear and feel. Can you give an example of how humans typically fail when estimating how long it takes light to travel to the moon?

35.4.9

Question: When we accept bodily sensation and emotions as key factor in human decisions like in moral condemnations; how would AI manage to assist in jurisdiction?

Next Generation Applications of AI

Chapter **36**

36.1 Next Generation Applications of Artificial Intelligence

36.1.1

[Next Generation Applications of Artificial Intelligence](#)

According to [Wikipedia](#): “ .. Artificial intelligence (AI) has been used in applications to alleviate certain problems throughout industry and academia. AI, like electricity or computers, is a general-purpose technology that has a multitude of applications. It has been used in fields of language translation, image recognition, credit scoring, e-commerce, and other domains. Through the Wikipedia link in the title of this paragraph, a detailed overview of the whole spectrum of AI applications can be retrieved. Please ask yourself which application strand you would never associate with AI and take the underlying question in mind before you start reading.”

36.1.2

Question: If you go along all mentioned AI application domains above; which of them would you favour to work on in further study projects and in your later professional career?

36.2 Artificial Intelligence in Industry

36.2.1

Artificial Intelligence in Industry



“ In this very special conversation, we had with us Anshuman Singh, the co-founder of Scaler. Scaler Academy in India is an online tech-‘varsity’ for the top 1% of software developers in the country. They offer an intensive six-month computer science course through live classes delivered by tech leaders and subject matter experts. Their meticulously structured program enhances the skills of software professionals by offering a modern curriculum with exposure to the latest technologies. In this clip, we discuss the correct way to study Artificial Intelligence and how you can become an artificial intelligence engineer.”

According to [Wikipedia](#): “ .. Industrial artificial intelligence, or industrial AI, usually refers to the application of artificial intelligence to industry. Unlike general artificial intelligence which is a frontier research discipline to build computerized systems that perform tasks requiring human intelligence, industrial AI is more concerned with the application of such technologies to address industrial pain-points for customer value creation, productivity improvement, cost reduction, site optimization, predictive analysis, and insight discovery. Although in a dystopian vision of AI applications, intelligent machines may take away jobs of humans and cause social and ethical

issues, industry in general holds a more positive view of AI and sees this transformation of economy unstoppable and expects huge business opportunities in this process. The concept of artificial intelligence was initially proposed in the 1940s, and the idea of improving productivity and gaining insights through smart analytics and modelling is not new. Artificial Intelligence and Knowledge-Based systems have been an active research branch of artificial intelligence for the entire product life cycle for product design, production planning, distribution, and field services. E-manufacturing systems and e-factories did not use the term “AI,” but they scale up modelling of engineering systems to enable complete integration of elements in the manufacturing eco-system for smart operation management. Recently, to accelerate leadership in AI initiative, the US government launched an official website www.ai.gov to highlight its priorities in the ai space. there are several reasons for the recent popularity of industrial AI: More affordable sensors and the automated process of data acquisition; More powerful computation capability of computers to perform more complex tasks at a faster speed with lower cost; Faster connectivity infrastructure and more accessible cloud services for data management and computing power outsourcing.

Categories

Technology alone never creates any business value if the problems in industry are not well studied. The major categories which industrial AI may contribute to include: product and service innovation, process improvement, and insight discovery. Cloud Foundry service platforms widely embed the artificial intelligent technologies. Cyber manufacturing systems also apply predictive analytics and cyber-physical modelling to address the gap between production and machine health for optimized productivity. Product applications for user value creation: Industrial AI can be embedded to existing products or services to make them more effective, reliable, safer, and to enhance their longevity. The automotive industry, for example, uses computer vision to avoid accidents and enable vehicles to stay in lane, facilitating safer driving.

In manufacturing, one example is the prediction of blade life for self-aware band saw machines, so that users will be able to rely on evidence of degradation rather than experience, which is safer, will extend blade life, and build up blade usage profile to help blade selection.

36.2.2

Question: Check the meaning of ‘[Cloud Foundry Service Platforms](#)’ and ask yourself how AI has the options to care and purify vital data?

36.2.3

Process applications for **productivity improvement**: Automation is one of the major aspects in process applications of industrial AI. With the help of AI, the scope and pace of automation have been fundamentally changed. AI technologies boost the performance and expand the capability of conventional AI applications. An example

is the collaborative robots. Collaborative robotic arms are able to learn the motion and path demonstrated by human operators and perform the same task. AI also automates the process that used to require human participation. An example is the Hong Kong subway, where an AI program decides the distribution and job scheduling of engineers with more efficiency and reliability than human counterparts do.

36.2.4

Question: The notion of collaborative robots have been coined. The human bodily interactions in for instance dancing might be captured for later artistic AI. What future applications can you predict?

36.2.5

Another aspect of process applications is the modelling large-scale systems. Cyber manufacturing systems are defined as a manufacturing service system that is networked and resilient to faults by evidence-based modelling and data-driven deep learning. Such a system deals with large and usually geographically distributed assets, which is hard to be modelled via conventional individual-asset physics-based model. With machine learning and optimization algorithms, a bottom-up framework considering machine health can leverage large samples of assets and automate the operation management, spare part inventory planning, and maintenance scheduling process.

36.2.6

Question and for Remembering. geographically distributed assets have been mentioned in the context of a bottom-up approach. How do you imagine its functioning for spare part inventory planning?

36.2.7

Insight applications for **knowledge discovery**: Industrial AI can also be used for knowledge discovery by identifying insights in engineering systems. In aviation and aeronautics, AI has been playing a vital role in many critical areas, one of which is safety assurance and root cause. NASA is trying to proactively manage risks to aircraft safety by analysing flight numeric data and text reports in parallel to not only detect anomalies but also relate it to the causal factors. This mined insight of why certain faults happen in the past will shed light on predictions of similar incidents in the future and prevent problems before they occur. Predictive and preventive maintenance through data-driven machine learning is also critical in cost reduction for industrial applications. Prognostics and health management (PHM) programs capture the opportunities at the shop floor by modelling equipment health degradation.

 36.2.8

Question: Knowledge discovery has been highlighted as we mentioned Alpha-Go. What message from the successes in Alpha-Go can go to AI-based Health Care?

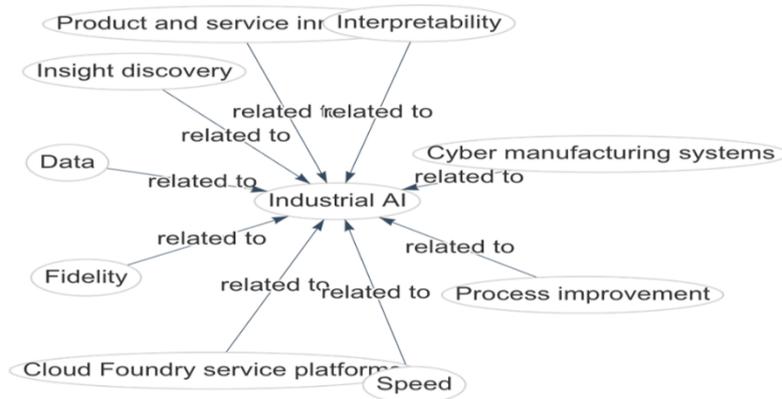
 36.2.9**Challenges**

The challenges of industrial AI to unlock the value lies in the transformation of raw data to intelligent predictions for rapid decision-making. In general, there are four major challenges in realizing industrial AI: data, speed, fidelity, and interpretability. Engineering systems now generate a lot of data and modern industry is indeed a big data environment. However, industrial data usually is structured, but may be low-quality. Production processes happen fast, and the equipment and work piece can be expensive. Therefore, the AI applications need to be applied in real-time to be able to detect anomalies immediately to avoid waste and other consequences. Cloud-based solutions can be powerful and fast, but they still would not fit certain computation efficiency requirements. Edge computing may be a better choice in such scenario. Unlike consumer-faced AI recommendations systems have a high tolerance for false positives and negatives. Even a very low rate of false positives or negatives rate may cost the total credibility of AI systems. Industrial AI applications are usually dealing with critical issues related to safety, reliability, and operations. Any failure in predictions could incur a negative economic and/or safety impact on the users and discourage them to rely on AI systems. Besides prediction accuracy and performance fidelity, the industrial AI systems must also go beyond prediction results and give root cause analysis for anomalies. This requires that during development, data scientists need to work with domain experts and include domain know-how into the modelling process, and have the model adaptively learn and accumulate such insights as knowledge.

 36.2.10

Question: The Wikipedia article above asserts that there are four major challenges in realizing industrial AI: data-driven, its speed, fidelity, and interpretability. The neural networking approach would complicate the criterion of interpretability. Could you describe in what sense interpretability will easily show its urgency in optimizing industrial processes; in particular at the aspect of fault finding.

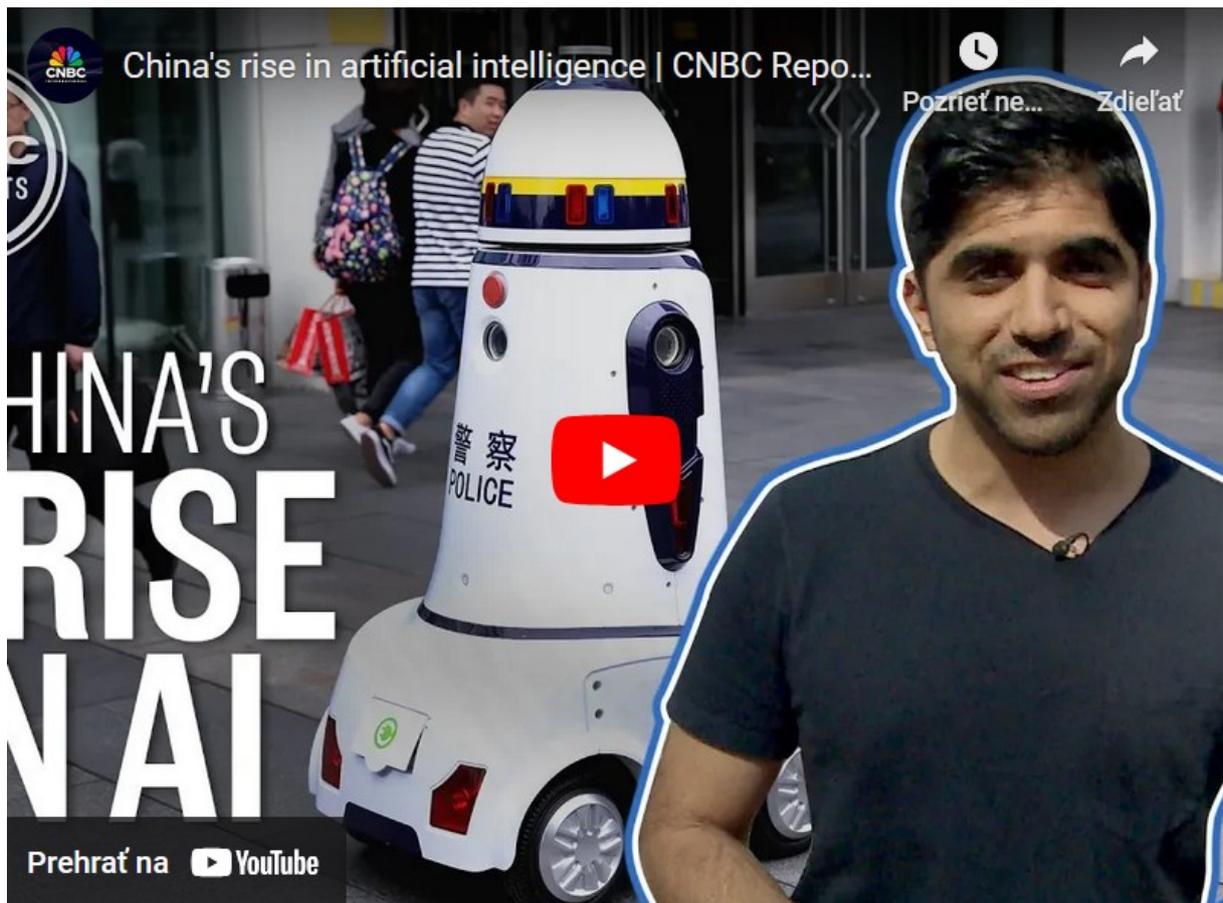
How far are we from Industrial A.I.?



36.3 Artificial Intelligence Industry in China

 36.3.1

[Artificial Intelligence Industry in China](#)



“ China's vast amount of data is helping advance its artificial intelligence research. CNBC's Uptin Saiidi speaks to experts VC Kai-Fu Lee, author Thomas Friedman and VC Benjamin Harburg on China's ambitions.”

36.3.2

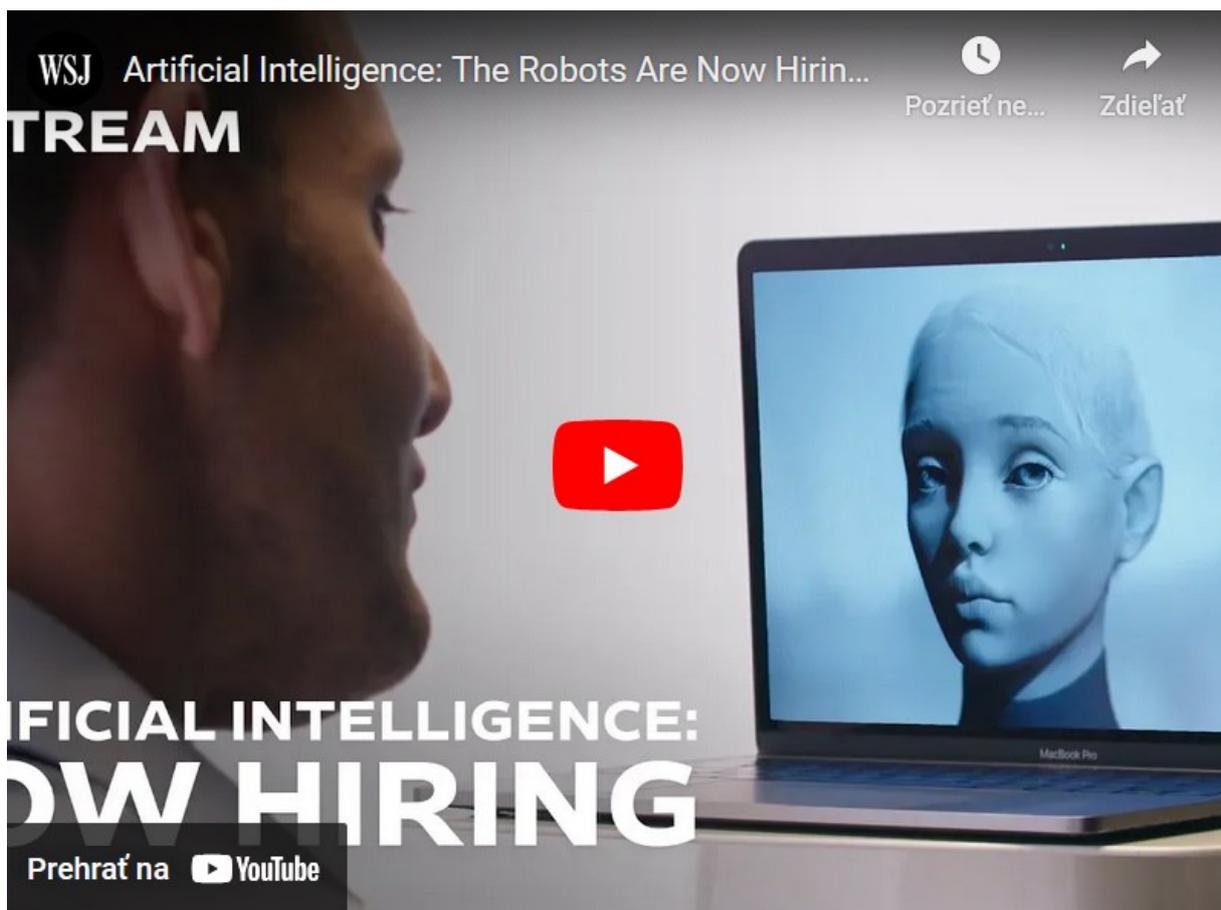
The artificial intelligence (AI) industry in China is a rapidly developing multi-billion-dollar industry. Please check the Wikipedia article as referred to in the linked title of this paragraph. After you read the Wikipedia article in the AI Industry in China, please try to answer the next question:

Recently (mid 2022) the geopolitical forces tend to fuel the alertness to China's strategic need for AI in mastering nano-meter granularity of chip design. AI and micro-chip seems mutually dependent. How would cognitive representations like cortical interconnectivity decide upon the next generation of computers for self-driving cars for instance? What ethical constraints would you see for excluding China from this techno-race?

36.4 Artificial Intelligence in Hiring

36.4.1

[Artificial Intelligence in Hiring](#)



“ Some Fortune 500 companies are using tools that deploy artificial intelligence to weed out job applicants. But is this practice fair? In this episode of Moving Upstream, WSJ's Jason Bellini investigates.”

Artificial intelligence (AI) in hiring involves the use of technology to automate aspects of the hiring process. Advances in artificial intelligence, such as the advent of machine learning and the growth of big data, enable AI to be utilized to recruit, screen, and predict the success of applicants. Proponents of artificial intelligence in hiring claim it reduces bias, assists with finding qualified candidates, and frees up human resource workers' time for other tasks, while opponents worry that AI perpetuates inequalities in the workplace and will eliminate jobs.

36.4.2

Question: Scouting talents for the labour market is often seen as a case of matching CVs with job demands. A more fundamental question is if head-hunters should stay at the level of match making. Scouting and finding unheard talents will keep new colleagues at the level of the existing ones. Steve Jobs complained ironically: “Apple scouts the best young talents world-wide, and pays them top salaries. Subsequently we tell them what to do ...”. Please reflect on this dilemma: What would be the best ethical way to attract young talents without narrowing the enterprise’s scope?

Regulation of AI

Chapter **37**

37.1 Regulation of Artificial Intelligence

37.1.1

Regulation of Artificial Intelligence



“ Please join the CSIS AI Governance Project for a fireside chat with Mr. Dragos Tudorache, Member of the European Parliament and Co-Rapporteur of the EU AI Act. The European Union is in the middle of finalizing the EU Artificial Intelligence Act, a globally first-of-its-kind attempt at horizontal AI regulation. Through the EU AI Act, the bloc seeks to act as a global standard setter on AI policy, creating a range of implications for the development of global AI regulation, as well as efforts to build international cooperation on AI. As one of the leading figures in Brussel's attempt to regulate artificial intelligence, MEP Tudorache also chairs the EU Special Committee on Artificial Intelligence in the Digital Age (AIDA). He will be joining CSIS to discuss the EU's approach to regulating AI, international collaboration on AI regulation, and key provisions within the draft EU AI Act.”

According to [Wikipedia](#): “ .. The regulation of artificial intelligence is the development of public sector policies and laws for promoting and regulating artificial intelligence (AI); it is therefore related to the broader regulation of algorithms. The regulatory and policy landscape for AI is an emerging issue in jurisdictions globally, including in the European Union and in supra-national bodies like the IEEE, OECD and others. Since

2016, a wave of AI ethics guidelines have been published in order to maintain social control over the technology. Regulation is considered necessary to both encourage AI and manage associated risks.”

37.1.2

Question: As AI research already reached the delicate level of weapon industry; can we still allow AI architectures to arrive in education and scientific publications?

37.2 Philosophy of Artificial Intelligence

37.2.1

[Philosophy of Artificial Intelligence](#)

Video is not available.

“ In this lesson we're going to introduce the subject of Artificial intelligence as a philosophical topic. In this first lesson we outline and explore the details of the course: How to Argue, Introduction to Philosophical Arguments. The Philosophy Academy is an educational project designed to teach Philosophical content for all, for free. Tags: philosophy, logic, education, philosophy lesson, lesson, teaching, online, Plato, university, free education, free philosophy, metaphysics, politics, political philosophy, ethics, language, history of philosophy, the learning academy, a level philosophy, how to argue, philosophical arguments, inductive arguments, deductive arguments, valid argument, sound arguments, fallacies, sentences and propositions, AI, artificial intelligence, Chinese room argument, computers, thinking computers.”

According to [Wikipedia](#): “ .. The philosophy of artificial intelligence is a branch of the philosophy of technology that explores artificial intelligence and its implications for knowledge and understanding of intelligence, ethics, consciousness, epistemology, and free will. Furthermore, the technology is concerned with the creation of artificial animals or artificial people (or, at least, artificial creatures; see artificial life) so the discipline is of considerable interest to philosophers. These factors contributed to the emergence of the philosophy of artificial intelligence. Some scholars argue that the AI community's dismissal of philosophy is detrimental. The philosophy of artificial intelligence attempts to answer such questions as follows:

- Can a machine act intelligently?
- Can it solve any problem that a person would solve by thinking?
- Are human intelligence and machine intelligence the same? Is the human brain essentially a computer?
- Can a machine have a mind, mental states, and consciousness in the same sense that a human being can?
- Can it feel how things are?

 37.2.2

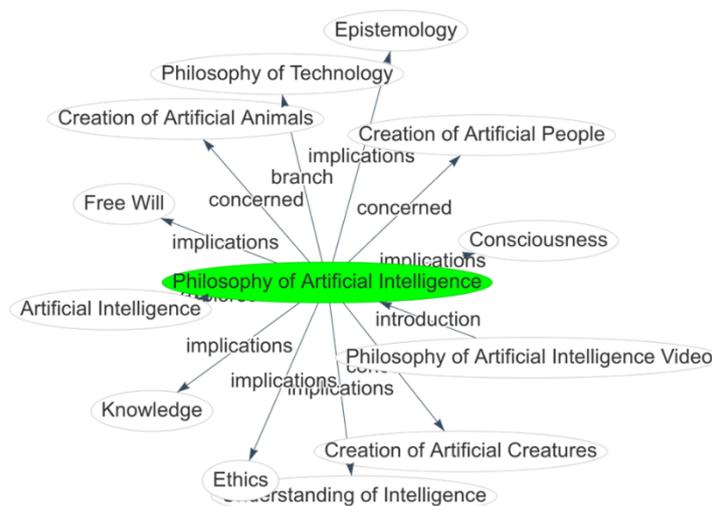
Question: Seen the evolutionary nature of human intellect; how can further AI development learn from animal to human cerebral development?

 37.2.3

Questions like these reflect the divergent interests of AI researchers, cognitive scientists and philosophers respectively. The scientific answers to these questions depend on the definition of "intelligence" and "consciousness" and exactly which "machines" are under discussion. Important propositions in the philosophy of AI include some of Turing's "polite convention": If a machine behaves as intelligently as a human being, then it is as intelligent as a human being.

Please check further the Wikipedia page as referred to above. In terms of the practical question of AI ("Can a machine display general intelligence?"), some versions of computationalism make the claim that (as Hobbes wrote): Reasoning is nothing but reckoning. In other words, our intelligence derives from a form of calculation, similar to arithmetic. This is the physical symbol system hypothesis discussed above, and it implies that artificial intelligence is possible. In terms of the philosophical question of AI ("Can a machine have a mind, mental states and consciousness?"), most versions of computationalism claim that (as Stevan Harnad characterizes it): Mental states are just implementations of (the right) computer programs. This is John Searle's "strong AI" discussed above, and it is the real target of the Chinese room argument. "

Philosophy of A.I.: do we Need to Revise Epistemics?



 37.2.4

Question: Turing's "polite convention": If a machine behaves as intelligently as a human being, then it is as intelligent as a human being. What abstraction legitimates Turing's thesis? In other words: what flaw can be seen in Turing's argumentation?

37.3 AI Ethics Fields' Approaches

37.3.1

[AI Ethics Fields' Approaches](#)



“ On a mission to drastically improve the concussion recovery process. Our brains are our greatest asset, and we should do everything we can to protect and heal them. With the use of machine learning we can gain valuable insights that can guide us throughout the recovery process. Ashleigh shares how to spark social engagement in the health care AI race, by encouraging us to embrace AI and take control of our health data. Ashleigh is a health technology entrepreneur from Stanford University with a PhD in Exercise Physiology and a post-doctoral fellowship in Neuroscience from the Toronto Rehabilitation Institute. She is the founder and CEO of Neurovine, a deep tech company delivering data driven brain health solutions. Before founding Neurovine, Ashleigh was a wearable technology consultant who worked with innovative start-ups and large-scale medical institutions. These experiences revealed the amazing potential of AI in health care but also highlighted the dangers of automation in this industry. This talk was given at a TEDx event using the TED conference format but independently organized by a local community.”

According to [Wikipedia](#): “ .. The ethics of artificial intelligence is the branch of the ethics of technology specific to artificially intelligent systems. It is sometimes

divided into a concern with the moral behaviour of humans as they design, make, use and treat artificially intelligent systems, and a concern with the behaviour of machines, in machine ethics. It also includes the issue of a possible singularity due to superintelligent AI. The term "robot ethics" (sometimes "roboethics") refers to the morality of how humans design, construct, use and treat robots. Robot ethics intersect with the ethics of AI. Robots are physical machines whereas AI can be only software. Not all robots function through AI systems and not all AI systems are robots. Robot ethics considers how machines may be used to harm or benefit humans, their impact on individual autonomy, and their effects on social justice. Machine ethics: Machine ethics (or machine morality) is the field of research concerned with designing Artificial Moral Agents (AMAs), robots or artificially intelligent computers that behave morally or as though moral. To account for the nature of these agents, it has been suggested to consider certain philosophical ideas, like the standard characterizations of agency, rational agency, moral agency, and artificial agency, which are related to the concept of AMAs. Please check further the work by Isaac Asimov in the 1950s in his "I, Robot". "

37.3.2

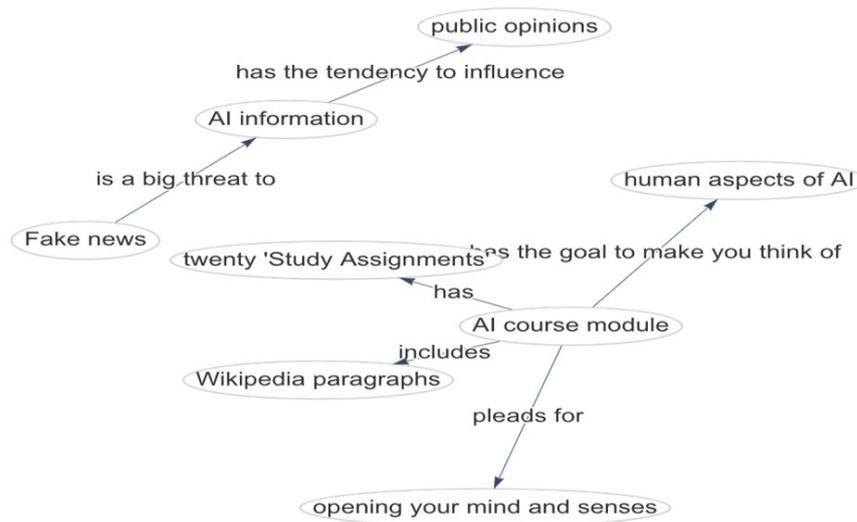
Questions: Most technological inventions in the last five centuries have triggered the question if not too close the human soul would be touched:

1. The printing press, that would bypass human handwriting and spread the word to those who were never involved in the bible before. Initially, the reprinted books caught Suspicion that God might become irritated as the Holy Word was conveyed without the human voice. Remember that "Per Sona" became later the equivalent of "Person". In other words: the 'frozen' word became portable and detached from the person-to-person interaction. Later, the same happened to the strip book, where chunks of text were captured in drawings; also here the irrevocable unity between the spoken and the printed word became broken and caused significant social resistance.
2. The entry of telephony, radio and television polarized society, even for religious reasons; "who had the right to counterfeit the human senses?" In large parts of Western society, the new media were seen as sabotage and cheating the unity of persons' trust as it was no longer sure if exactly these words or human gesture were uttered.
3. Quite recently, the social media, fake news and the deep fake entered our daily mental 'territories' and are seen as potentially devastating for mutual trust and social cohesion.
4. AI as driver behind deep fake and state propaganda has a similar threat to the commons opinion; similar to the magnitude of societal indignation in all three earlier media transitions as just listed before.

37.3.3

Please synthesize all you just learnt before, on the AI impact to all facets of human culture and try to predict what time it will still cost to let society exploit and trust AI as tool for making life easier, more reliable and finally more transparent as well.

Facing Twenty Study Assignments



37.4 EU AI Act - first regulation on AI

📖 37.4.1

Artificial intelligence holds immense potential to improve many aspects of our lives, from healthcare to transport to energy efficiency. However, as with any powerful technology, it can also pose risks if not carefully managed. The European Union recognized the need to regulate AI to maximize its benefits while minimizing potential harms. One of the primary reasons for regulation is to ensure that AI systems are safe and reliable for users. AI systems that make decisions or control critical infrastructure, like transportation or healthcare, can significantly affect people's safety and well-being. Regulation helps to ensure these systems are designed, tested, and monitored to prevent harm.

Regulation is not only about preventing risks but also about establishing ethical guidelines for the development and use of AI. The EU, for instance, prioritizes creating transparent, non-discriminatory, and environmentally friendly AI systems. This means that AI should be designed to be understandable to users, not operate in a "black box" manner, and avoid biases that might harm individuals or groups. Furthermore, AI systems should be used responsibly, with human oversight to prevent harmful consequences such as job displacement or discrimination.

EU AI act:

<https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>

 37.4.2

Which of the following is NOT a primary goal of AI regulation in the EU?

- Making AI systems secret and non-transparent
- Preventing harm by establishing ethical guidelines
- Ensuring AI systems are safe and reliable for users
- Promoting non-discriminatory and environmentally friendly AI systems

 37.4.3

Risk level

The EU's AI regulation framework classifies AI systems into different categories based on the level of risk they pose to users. These categories help determine how much regulation each AI system requires. AI systems with **unacceptable risk** - such as those that manipulate people's behavior or use facial recognition for surveillance - are banned entirely. This strict approach ensures that AI technologies are not used to infringe on people's privacy, rights, or safety.

On the other hand, **high-risk** AI systems - like those used in healthcare, transportation, or law enforcement - require extensive oversight and must meet specific safety and ethical standards before being deployed. These systems are evaluated not only at the time of their release but also throughout their lifecycle to ensure they continue to meet safety standards. By categorizing AI systems by risk, the EU can tailor the regulations to the level of threat each system poses, ensuring that high-risk systems receive the attention they need while less risky systems are still monitored effectively.

 37.4.4

Which of the following is classified as an **unacceptable risk** AI system?

- AI systems used in facial recognition for surveillance
- AI in transportation to improve vehicle safety
- AI used for managing critical infrastructure
- AI used in healthcare to diagnose diseases

 37.4.5

Transparency

Transparency is another key element in AI regulation. AI systems, especially those that generate content like ChatGPT, must be designed to disclose that they are AI-generated. This ensures that users understand when they are interacting with a machine rather than a human. It is also important that AI systems are designed to prevent the generation of illegal content, such as hate speech or false information.

These transparency requirements are not only for ethical reasons but also to maintain trust in AI technologies and to ensure that users are fully aware of the AI's capabilities and limitations.

Transparency extends beyond content generation. It also includes ensuring that AI systems can be traced and understood, which is essential for accountability. For example, if an AI system makes a harmful decision, it is important to understand how it came to that conclusion. In this way, transparency helps ensure that AI systems operate in a manner that is ethical, legal, and aligned with the values of society.

37.4.6

Which of the following is a requirement for AI systems in terms of transparency?

- Content generated by AI should be disclosed to users as AI-generated
- AI systems should be used for surveillance purposes without explanation
- AI systems should operate without any human oversight
- AI systems should remain opaque and untraceable

37.4.7

Balancing innovation with regulation

While regulation is essential for ensuring safety and ethics, it is also important to support innovation in AI development. The EU's AI regulation framework strikes a balance by offering start-ups and small businesses opportunities to develop AI technologies in a controlled environment. This is done through testing environments that simulate real-world conditions, allowing companies to refine their products before they are released to the public. These testing environments help to ensure that AI systems meet regulatory standards while giving companies the space to innovate and grow.

In addition, the regulation framework provides specific timelines for AI companies to comply with the requirements, giving them time to adjust their systems. For example, high-risk AI systems will need to meet compliance standards within three years, while other systems will be subject to requirements sooner. This phased approach allows the AI industry to adapt while maintaining a focus on safety, ethics, and transparency.

37.4.8

Which of the following is a way the EU encourages innovation in AI?

- By allowing testing environments for start-ups and small businesses
- By immediately implementing strict regulations on all AI systems
- By restricting all AI research to only large companies
- By banning AI systems that pose any risk, regardless of their potential benefits

 37.4.9

Can you Remember? What are the reasons of AI regulations?

 37.4.10

Question: Given the increasing importance of AI regulation in ensuring the ethical use of technology, especially with regard to transparency, risk categorization, and innovation, how do you think the education system should adapt its curriculum and instructional methods to better prepare students for understanding and engaging with AI regulation and its implications in various sectors?

Additional Resources for Further Study

Chapter **38**

38.1 Additional resources for further Study I.

38.1.1

Thamik, H.; Wu, J. (2022) The Impact of Artificial Intelligence on Sustainable Development in Electronic Markets. Sustainability 2022, 14, 3568. <https://doi.org/10.3390/su14063568>

With the emergence of artificial intelligence (AI), the technological revolution has transformed human lives and processes, empowering the products and services in today's marketplaces. AI introduces new ways of doing jobs and business, and of exploring new global market opportunities. However, on the other hand, it provides many challenges to comprehend. Therefore, our study's main objective was to examine the behavioral, cultural, ethical, social, and economic challenges of AI-enabled products and services in consumer markets and discuss how businesses might shape their approaches to address AI-related ethical issues. AI offers numerous social, ethical, and behavioral difficulties for people, organizations, and societies that endanger the sustainable development of economies.

These fundamental challenges due to AI technologies have raised serious questions for the sustainable development of electronic markets. Based on this, the current study presents a framework highlighting these issues. Systematic reviewing was our research method; we looked for explicit information and methods that indicate the credibility of research or reduce biases. This paper is of great importance, as it highlights several behavioral, societal, ethical, and cultural aspects in electronic markets which were not presented in previous studies. Some key issues are the security and privacy of consumers, AI biases, individual autonomy, wellbeing, and issues of unemployment.

Therefore, companies that use AI systems need to be socially responsible and make AI systems as secure as possible to promote the sustainable development of countries. The results suggest that AI has undoubtedly transformed life and has both positive and negative effects. However, the main aim regarding AI should be to use it for the overall goals of humanity. Moreover, authorities operating in e-business environments need to create proper rules and regulations and make the systems as secure as possible for people.

Why companies that use AI systems need to be socially responsible and make AI systems as secure as possible?

38.1.2

Malyshkin, A.V. (2019). Integrating Artificial Intelligence into Public Life: Some Ethical and Legal Issues. Herald St. Petersburg University. Law 3: 444–460. <https://doi.org/10.21638/spbu14.2019.303>

The spread of artificial intelligence systems gives rise to a number of technical, philosophical, legal and ethical issues related both to the admissibility of using such systems in certain areas, and the need to comply with ethical standards when creating them, as well as the possibility of introducing ethical standards into the decision-making process artificially. intellect.

Since for many people religion is the basis of the worldview, being present in public life precisely as ethics, and not dogmatics, the study of various aspects of the relationship between religion and artificial intelligence is also extremely relevant. The author analyzes the ethical and religious problems associated with the creation and dissemination of artificial intelligence systems, and suggests ways of legal regulation of social relations associated with the use of artificial intelligence.

In particular, the following issues are touched upon: legal personality and delinquency of artificial intelligence; Social and Ethical Aspects of Artificial Intelligence 57 sacralization of artificial intelligence; liability for harm caused by artificial intelligence systems; decision-making by artificial intelligence regarding the rights and obligations of people; exacerbation of stratification and inequality; mass unemployment; intellectual superiority of carriers of artificial intelligence over humans; alienation of people from each other, human loneliness; the possibility of following ethical standards when making decisions by artificial intelligence. (Malyshkin, 2019).

Why decision-making by artificial intelligence regarding the rights and obligations of people is important to consider and deeper analysis?

38.1.3

Stahl, B.C., Rodrigues, R., Santiago, N., Macnish, K. (2022) A European Agency for Artificial Intelligence: Protecting fundamental rights and ethical values. 45, Computer Law and Security Review ISSN 0267-3649, <https://doi.org/10.1016/j.clsr.2022.105661>.

Following years of intensive international debate of the ethical and human rights implications of artificial intelligence (AI)-related technologies, there are numerous proposals to legislate and regulate these technologies. One aspect of possible legislative frameworks for AI is the creation of a new regulator or other body with the remit to provide oversight of AI.

This article reviews the ethical and human rights challenges as well as proposed mitigation strategies, in order to discuss how a regulatory body might be designed to address these challenges. It focuses on a particular form that a new

body might take, more specifically on a potential European Agency for AI. Based on a multi-step methodology of stakeholder interaction, the article proposes a terms of reference for such an Agency and discusses the characteristics it would need to display to ensure that it could adequately engage with current and future ethical and human rights challenges arising from the development, deployment and use of AI. This proposal is then contrasted with the proposed European Artificial Intelligence Board included in the draft European Regulation on AI (the AI Act). (Stahl, B.C., Rodrigues, R., Santiago, N., Macnish, K., 2022)

Why the european agency for ai should be start and in which aims and tasks it should engage?

38.1.4

Karnouskos, S. (2022) Symbiosis with artificial intelligence via the prism of law, robots, and society. Artif Intell Law 30, 93–115 (2022). <https://doi.org/10.1007/s10506-021-09289-1>

The rapid advances in Artificial Intelligence and Robotics will have a profound impact on society as they will interfere with the people and their interactions. Intelligent autonomous robots, independent if they are humanoid/anthropomorphic or not, will have a physical presence, make autonomous decisions, and interact with all stakeholders in the society, in yet unforeseen manners. The symbiosis with such sophisticated robots may lead to a fundamental civilizational shift, with far-reaching effects as philosophical, legal, and societal questions on consciousness, citizenship, rights, and legal entity of robots are raised.

The aim of this work is to understand the broad scope of potential issues pertaining to law and society through the investigation of the interplay of law, robots, and society via different angles such as law, social, economic, gender, and ethical perspectives. The results make it evident that in an era of symbiosis with intelligent autonomous robots, the law systems, as well as society, are not prepared for their prevalence. Therefore, it is now the time to start a multi-disciplinary stakeholder discussion and derive the necessary policies, frameworks, and roadmaps for the most eminent issues. Karnouskos, S. (2022)

Why is it time to start a multidisciplinary discussion with stakeholders and develop the necessary policies, frameworks and action plans for the most social important issues, related to a wide range of potential problems related to law and society by examining the interplay of law, robots and society from different angles such as legal, social, economic, gender and ethical perspectives.

38.1.5

Ziosi, M., Hewitt, B., Juneja, P. et al. (2022) Smart cities: reviewing the debate about their ethical implications. *AI & Soc* (2022). <https://doi.org/10.1007/s00146-022-01558-0>

This paper considers a host of definitions and labels attached to the concept of smart cities to identify four dimensions that ground a review of ethical concerns emerging from the current debate. These are: (1) network infrastructure, with the corresponding concerns of control, surveillance, and data privacy and ownership; (2) post-political governance, embodied in the tensions between public and private decision-making and cities as post-political entities; (3) social inclusion, expressed in the aspects of citizen participation and inclusion, and inequality and discrimination; and (4) sustainability, with a specific focus on the environment as an element to protect but also as a strategic element for the future.

Given the persisting disagreements around the definition of a smart city, the article identifies in these four dimensions a more stable reference framework within which ethical concerns can be clustered and discussed. Identifying these dimensions makes possible a review of the ethical implications of smart cities that is transversal to their different types and resilient towards the unsettled debate over their definition. (Ziosi, M., Hewitt, B., Juneja, P. et al., 2022)

Please, identify four dimensions that could ground a review of ethical concerns emerging from the debate on developing of the smart cities.

38.1.6

Cox, A. (2022) The Ethics of AI for Information Professionals: Eight Scenarios, *Journal of the Australian Library and Information Association*, 71:3, 201-214, DOI: 10.1080/24750158.2022.2084885

Artificial Intelligence (AI) is central to transformative changes happening in many industries, perhaps potentially to a fourth industrial revolution, but it has also raised a storm of ethical concerns. Information professionals need to navigate these ethical issues effectively because they are likely to use AI in delivering services as well as contributing to the process of adoption of AI more widely in their organisations. Professional ethical codes are too high level to offer precise or complete guidance. In this context, the purpose of this paper is to review the relevant literature and describe eight ethics scenarios of AI which have been developed specifically for information professionals to understand the issues in a concrete form.

The paper considers how AI might be defined and presents some of the applications relevant to the information profession. It then summarizes the key ethical issues raised by AI in general both those inherent to the technology and those arising from the nature of the AI industry. It considers existing studies that have discussed aspects of the ethical issues specifically for information professionals. It then

describes a set of eight ethics scenarios that have been developed and shared in an open form to promote their reuse. Cox, A. (2022)

Please, describe eight ethics scenarios of ai which have been developed specifically for information professionals to understand the issues.

38.1.7

Hagendorff, T. (2021) Linking Human and Machine Behavior: A New Approach to Evaluate Training Data Quality for Beneficial Machine Learning. *Minds & Machines* 31, 563–593 (2021). <https://doi.org/10.1007/s11023-021-09573-8>

Machine behavior that is based on learning algorithms can be significantly influenced by the exposure to data of different qualities. Up to now, those qualities are solely measured in technical terms, but not in ethical ones, despite the significant role of training and annotation data in supervised machine learning. This is the first study to fill this gap by describing new dimensions of data quality for supervised machine learning applications.

Based on the rationale that different social and psychological backgrounds of individuals correlate in practice with different modes of human–computer-interaction, the paper describes from an ethical perspective how varying qualities of behavioral data that individuals leave behind while using digital technologies have socially relevant ramifications for the development of machine learning applications. The specific objective of this study is to describe how training data can be selected according to ethical assessments of the behavior it originates from, establishing an innovative filter regime to transition from the big data rationale $n = \text{all}$ to a more selective way of processing data for training sets in machine learning. The overarching aim of this research is to promote methods for achieving beneficial machine learning applications that could be widely useful for industry as well as academia. Hagendorff, T. (2021)

List and describe methods for obtaining beneficial machine learning applications that can be widely useful in both industry and academia.

38.1.8

Yam, Josephine & Skorburg, Joshua August (2021). From human resources to human rights: Impact assessments for hiring algorithms. *Ethics and Information Technology* 23 (4):611-623.

Over the years, companies have adopted hiring algorithms because they promise wider job candidate pools, lower recruitment costs and less human bias. Despite these promises, they also bring perils. Using them can inflict unintentional harms on individual human rights. These include the five human rights to work, equality and nondiscrimination, privacy, free expression and free association. Despite the human rights harms of hiring algorithms, the AI ethics literature has predominantly focused on abstract ethical principles. This is problematic for two reasons. First, AI principles

have been criticized for being vague and not actionable. Second, the use of vague ethical principles to discuss algorithmic risks does not provide any accountability. This lack of accountability creates an algorithmic accountability gap.

Closing this gap is crucial because, without accountability, the use of hiring algorithms can lead to discrimination and unequal access to employment opportunities. This paper makes two contributions to the AI ethics literature. First, it frames the ethical risks of hiring algorithms using international human rights law as a universal standard for determining algorithmic accountability. Second, it evaluates four types of algorithmic impact assessments in terms of how effectively they address the five human rights of job applicants implicated in hiring algorithms. It determines which of the assessments can help companies audit their hiring algorithms and close the algorithmic accountability gap. (Yam, Josephine & Skorburg, Joshua August, 2021).

Please, frame the ethical risks of hiring algorithms using international human rights law as a universal standard for determining algorithmic accountability. Please, evaluate four types of algorithmic impact assessments in terms of how effectively they address the five human rights of job applicants implicated in hiring algorithms.

38.1.9

Júlia Pareto Boada, Begoña Román Maestre, Carme Torras Genís, (2021) The ethical issues of social assistive robotics: A critical literature review, *Technology in Society*, Volume 67, ISSN 0160-791X, <https://doi.org/10.1016/j.techsoc.2021.101726>.

Along with its potential contributions to the practice of care, social assistive robotics raises significant ethical issues. The growing development of this technoscientific field of intelligent robotics has thus triggered a widespread proliferation of ethical attention towards its disruptive potential. However, the current landscape of ethical debate is fragmented and conceptually disordered, endangering ethics' practical strength for normatively addressing these challenges.

This paper presents a critical literature review of the ethical issues of social assistive robotics, which provides a comprehensive and intelligible overview of the current ethical approach to this technoscientific field. On the one hand, ethical issues have been identified, quantitatively analyzed and categorized in three main thematic groups. Namely: Well-being, Care, and Justice. On the other hand –and on the basis of some significant disclosed tendencies of the current approach–, future lines of research and issues regarding the enrichment of the ethical gaze on social assistive robotics have been identified and outlined.

Please, identify ethical issues, quantitatively analyzed and categorized in three main thematic groups. Please, future lines of research and issues regarding the enrichment of the ethical gaze on social assistive robotics have been identify and outline.

38.1.10

Rakowski, R., Polak, P. & Kowalikova, P. (2021) Ethical Aspects of the Impact of AI: the Status of Humans in the Era of Artificial Intelligence. Soc 58, 196–203 (2021). <https://doi.org/10.1007/s12115-021-00586-8>

On the one hand, AI is a functional tool for emancipating people from routine work tasks, thus expanding the possibilities of their self-realization and the utilization of individual interests and aspirations through more meaningful spending of time. On the other hand, there are undisputable risks associated with excessive machine autonomy and limited human control, based on the insufficient ability to monitor the performance of these systems and to prevent errors or damage (Floridi et al. *Minds & Machines* 28, 689–707, 2018).

In connection with the use of ethical principles in the research and development of artificial intelligence, the question of the social control of science and technology opens out into an analysis of the opportunities and risks that technological progress can mean for security, democracy, environmental sustainability, social ties and community life, value systems, etc. For this reason, it is necessary to identify and analyze the aspects of artificial intelligence that could have the most significant impact on society. The present text is focused on the application of artificial intelligence in the context of the market and service sector, and the related process of exclusion of people from the development, production and distribution of goods and services. Should the application of artificial intelligence be subject to value frameworks, or can the application of AI be sufficiently regulated by the market on its own?

Should the application of artificial intelligence be subject to value frameworks, or can the application of ai be sufficiently regulated by the market on its own?

38.2 Additional resources for further Study II.

38.2.1

Han Yu, Zhiqi Shen, Chunyan Miao, Cyril Leung, Victor R., & Qiang Yang (2018) Lesser Building Ethics into Artificial Intelligence Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)

As artificial intelligence (AI) systems become increasingly ubiquitous, the topic of AI governance for ethical decision-making by AI has captured public imagination. Within the AI research community, this topic remains less familiar to many researchers. In this paper, we complement existing surveys, which largely focused on the psychological, social and legal discussions of the topic, with an analysis of recent advances in technical solutions for AI governance.

By reviewing publications in leading AI conferences including AAAI, AAMAS, ECAI and IJCAI, we propose a taxonomy which divides the field into four areas: 1) exploring

ethical dilemmas; 2) individual ethical decision frameworks; 3) collective ethical decision frameworks; and 4) ethics in human-AI interactions. We highlight the intuitions and key techniques used in each approach, and discuss promising future research directions towards successful integration of ethical AI systems into human societies.

Please, analyse a taxonomy which divides the field of on the psychological, social and legal discussions for ai governance into four areas: 1) exploring ethical dilemmas; 2) individual ethical decision frameworks; 3) collective ethical decision frameworks; and 4) ethics in human-ai interactions.

38.2.2

Cath, C. (2018) Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376(2133), <https://doi.org/10.1098/rsta.2018.0080>

This paper is the introduction to the special issue entitled: 'Governing artificial intelligence: ethical, legal and technical opportunities and challenges'. Artificial intelligence (AI) increasingly permeates every aspect of our society, from the critical, like urban infrastructure, law enforcement, banking, healthcare and humanitarian aid, to the mundane like dating. AI, including embodied AI in robotics and techniques like machine learning, can improve economic, social welfare and the exercise of human rights. Owing to the proliferation of AI in high-risk areas, the pressure is mounting to design and govern AI to be accountable, fair and transparent.

How can this be achieved and through which frameworks? This is one of the central questions addressed in this special issue, in which eight authors present in-depth analyses of the ethical, legal-regulatory and technical challenges posed by developing governance regimes for AI systems. It also gives a brief overview of recent developments in AI governance, how much of the agenda for defining AI regulation, ethical frameworks and technical approaches is set, as well as providing some concrete suggestions to further the debate on AI governance. This article is part of the theme issue 'Governing artificial intelligence: ethical, legal, and technical opportunities and challenges'. Cath, C. (2018)

How can be achieved and through which frameworks the aims, concerning regulation of using ai in robotics and techniques like machine learning, owing to the proliferation of ai in high-risk areas, and the pressure is mounting to design and govern ai to be accountable, fair and transparent?

 38.2.3

Hu, L. (2018) **Justice Beyond Utility in Artificial Intelligence**. AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. (pp. 368-369) <https://doi.org/10.1145/3278721.3278798>

The entry of Artificial Intelligence into prominent social and economic environments has brought to the fore concerns about the ethical nature of such agents and tools. Though it is well-known that methods based in utilitarian calculus often fail to account for moral considerations, few alternatives have been adopted in the field of AI.

This work advocates for a new approach toward the interaction between AI methods and the social that centres principles of distributive justice. As AI increasingly drives consequential social decision-making, we must consider not only what can be done but what ought to be done. Grappling with the inherently normative nature of these problems requires an orientation towards AI that is able to conceptualize justice beyond utility.

Please describe a new approach toward the interaction between ai methods and the social that centers principles of distributive justice.

 38.2.4

Latham, A., Goltz, S. (2019). **A Survey of the General Public's Views on the Ethics of Using AI in Education**. In: Isotani, S., Millán, E., Ogan, A., Hastings, P., McLaren, B., Luckin, R. (eds) **Artificial Intelligence in Education**. AIED 2019. Lecture Notes in Computer Science(), vol 11625. Springer, Cham. https://doi.org/10.1007/978-3-030-23204-7_17

Recent scandals arising from the use of algorithms for user profiling to further political and marketing gain have popularized the debate over the ethical and legal implications of using such 'artificial intelligence' in social media. The need for a legal framework to protect the general public's data is not new, yet it is not clear whether recent changes in data protection law in Europe, with the introduction of the GDPR, have highlighted the importance of privacy and led to a healthy concern from the general public over online user tracking and use of data. Like search engines, social media and online shopping platforms, intelligent tutoring systems aim to personalize learning and thus also rely on algorithms that automatically profile individual learner traits.

A number of studies have been published on user perceptions of trust in robots and computer agents. Unsurprisingly, studies of AI in education have focused on efficacy, so the extent of learner awareness, and acceptance, of tracking and profiling algorithms remains unexplored. This paper discusses the ethical and legal considerations for, and presents a case study examining the general public's views of, AI in education. A survey was recently taken of attendees at a national science festival event highlighting state-of-the-art AI technologies in education. Whilst most

participants (77%) were worried about the use of their data, in learning systems fewer than 8% of adults were 'not happy' being tracked, as opposed to nearly two-thirds (63%) of children surveyed.

Please presented your opinion on the ethical and legal considerations for, which were presented a case study examining the general public's views of, AI in education and some results of survey was recently taken of attendees at a national science festival event highlighting state-of-the-art AI technologies in education.

38.2.5

Ahmed Z, Mohamed K, Zeeshan S, Dong X. (2020) Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database (Oxford). 2020 Jan 1;2020:baaa010.1-35. <https://doi.org/10.1093/database/baaa010>

Precision medicine is one of the recent and powerful developments in medical care, which has the potential to improve the traditional symptom-driven practice of medicine, allowing earlier interventions using advanced diagnostics and tailoring better and economically personalized treatments. Identifying the best pathway to personalized and population medicine involves the ability to analyse comprehensive patient information together with broader aspects to monitor and distinguish between sick and relatively healthy people, which will lead to a better understanding of biological indicators that can signal shifts in health. While the complexities of disease at the individual level have made it difficult to utilize healthcare information in clinical decision-making, some of the existing constraints have been greatly minimized by technological advancements.

To implement effective precision medicine with enhanced ability to positively impact patient outcomes and provide real-time decision support, it is important to harness the power of electronic health records by integrating disparate data sources and discovering patient-specific patterns of disease progression. Useful analytic tools, technologies, databases, and approaches are required to augment networking and interoperability of clinical, laboratory and public health systems, as well as addressing ethical and social issues related to the privacy and protection of healthcare data with effective balance.

Developing multifunctional machine learning platforms for clinical data extraction, aggregation, management and analysis can support clinicians by efficiently stratifying subjects to understand specific scenarios and optimize decision-making. Implementation of artificial intelligence in healthcare is a compelling vision that has the potential in leading to the significant improvements for achieving the goals of providing real-time, better personalized and population medicine at lower costs. In this study, we focused on analysing and discussing various published artificial intelligence and machine learning solutions, approaches and perspectives, aiming to advance academic solutions in paving the way for a new data-centric era of discovery in healthcare.

What do you think about implementation of artificial intelligence in healthcare? Could this has the potential in leading to the significant improvements for achieving the goals of providing real-time, better personalized and population medicine at lower costs? How?

 38.2.6

Zhang, B., Dafoe, A. (2020) U.S. Public opinion on the governance of artificial intelligence. AIES 2020 - Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society U.S. Public opinion on the governance of artificial intelligence. (pp. 187-193) <https://doi.org/10.1145/3375627.3375827>

Artificial intelligence (AI) has widespread societal implications, yet social scientists are only beginning to study public attitudes toward the technology. Existing studies find that the public's trust in institutions can play a major role in shaping the regulation of emerging technologies. Using a large-scale survey (N=2000), we examined Americans' perceptions of 13 AI governance challenges as well as their trust in governmental, corporate, and multistakeholder institutions to responsibly develop and manage AI. While Americans perceive all of the AI governance issues to be important for tech companies and governments to manage, they have only low to moderate trust in these institutions to manage AI applications.

King, O.C. (2020) Presumptuous aim attribution, conformity, and the ethics of artificial social cognition. *Ethics Inf Technol* 22, 25–37 (2020). <https://doi.org/10.1007/s10676-019-09512-3>

Imagine you are casually browsing an online bookstore, looking for an interesting novel. Suppose the store predicts you will want to buy a particular novel: the one most chosen by people of your same age, gender, location, and occupational status. The store recommends the book, it appeals to you, and so you choose it. Central to this scenario is an automated prediction of what you desire. This article raises moral concerns about such predictions. More generally, this article examines the *ethics of artificial social cognition*—the ethical dimensions of attribution of mental states to humans by artificial systems.

The focus is *presumptuous* aim attributions, which are defined here as aim attributions based crucially on the premise that the person in question will have aims like superficially similar people. Several everyday examples demonstrate that this sort of presumptuousness is already a familiar moral concern.

The scope of this moral concern is extended by new technologies. In particular, recommender systems based on collaborative filtering are now commonly used to automatically recommend products and information to humans. Examination of these systems demonstrates that they naturally attribute aims presumptuously.

This article presents two reservations about the widespread adoption of such systems. First, the severity of our antecedent moral concern about presumptuousness increases when aim attribution processes are automated and accelerated. Second, a foreseeable consequence of reliance on these systems is an unwarranted inducement of interpersonal conformity.

Please describe two reservations about the widespread adoption of systems, which now commonly used to automatically recommend products and information to humans. First, the severity of our antecedent moral concern about presumptuousness increases when aim attribution processes are automated and accelerated. Second, a foreseeable consequence of reliance on these systems is an unwarranted inducement of interpersonal conformity.

38.2.7

Findlay, M., Seah, J. (2020) An Ecosystem Approach to Ethical AI and Data Use: Experimental Reflections 2020 IEEE / ITU International Conference on Artificial Intelligence for Good, AI4G 2020 9311069, pp. 192-197
<https://doi.org/10.48550/arXiv.2101.02008>

While we have witnessed a rapid growth of ethics documents meant to guide AI development, the promotion of AI ethics has nonetheless proceeded with little input from AI practitioners themselves. Given the proliferation of AI for Social Good initiatives, this is an emerging gap that needs to be addressed in order to develop more meaningful ethical approaches to AI use and development. This paper offers a methodology, a shared fairness approach, aimed at identifying the needs of AI practitioners when it comes to confronting and resolving ethical challenges and to find a third space where their operational language can be married with that of the more abstract principles that presently remain at the periphery of their work experiences. We offer a grassroots approach to operational ethics based on dialog and mutualized responsibility.

This methodology is centred around conversations intended to elicit practitioners perceived ethical attribution and distribution over key value laden operational decisions, to identify when these decisions arise and what ethical challenges they confront, and to engage in a language of ethics and responsibility which enables practitioners to internalize ethical responsibility. The methodology bridges responsibility imbalances that rest in structural decision making power and elite technical knowledge, by commencing with personal, facilitated conversations, returning the ethical discourse to those meant to give it meaning at the sharp end of the ecosystem. Our primary contribution is to add to the recent literature seeking to bring AI practitioners' experiences to the fore by offering a methodology for understanding how ethics manifests as a relational and interdependent sociotechnical practice in their work.

Please describe a methodology, a shared fairness approach, aimed at identifying the needs of ai practitioners when it comes to confronting and resolving ethical challenges and to find a third space where their operational language can be

married with that of the more abstract principles that presently remain at the periphery of their work experiences.

38.2.8

Muhlenbach, F. (2020) A Methodology for Ethics-by-Design AI Systems: Dealing with Human Value Conflicts. Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics. 2020-October, 9283185, pp. 1310-1315. <https://doi.org/10.48550/arXiv.2010.07610>

The introduction of artificial intelligence into activities traditionally carried out by human beings produces brutal changes. This is not without consequences for human values. This paper is about designing and implementing models of ethical behaviours in AI-based systems, and more specifically it presents a methodology for designing systems that take ethical aspects into account at an early stage while finding an innovative solution to prevent human values from being affected. Two case studies where AI-based innovations complement economic and social proposals with this methodology are presented: one in the field of culture and operated by a private company, the other in the field of scientific research and supported by a state organization.

Please present designing and implementing models of ethical behaviors in AI-based systems, and more specifically it presents a methodology for designing systems that take ethical aspects into account at an early stage while finding an innovative solution to prevent human values from being affected and describe two case studies where AI-based innovations complement economic and social proposals with this methodology.

38.2.9

Marek Havrda, Bogdana Rakova (2020) Enhanced well-being assessment as basis for the practical implementation of ethical and rights-based normative principles for AI. 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC) Oct 2020 Pages 2754–2761 <https://doi.org/10.1109/SMC42975.2020.9283137>

impact on all areas of people's livelihoods. A detailed look at existing interdisciplinary and transdisciplinary metrics frameworks could bring new insights and enable practitioners to navigate the challenge of understanding and assessing the impact of Autonomous and Intelligent Systems (A/IS). There has been emerging consensus on fundamental ethical and rights-based AI principles proposed by scholars, governments, civil rights organizations, and technology companies. In order to move from principles to real-world implementation, we adopt a lens motivated by regulatory impact assessments and the well-being movement in public policy. Similar to public policy interventions, outcomes of AI systems implementation may have far-reaching complex impacts. In public policy, indicators are only part of a broader toolbox, as metrics inherently lead to gaming and dissolution of incentives and objectives.

Similarly, in the case of A/IS, there is a need for a larger toolbox that allows for the iterative assessment of identified impacts, inclusion of new impacts in the analysis, and identification of emerging trade-offs. In this paper, we propose the practical application of an enhanced well-being impact assessment framework for A/IS that could be employed to address ethical and rights-based normative principles in AI. This process could enable a human-centered algorithmically supported approach to the understanding of the impacts of AI systems. Finally, we propose a new testing infrastructure which would allow for governments, civil rights organizations, and others, to engage in cooperating with A/IS developers towards implementation of enhanced well-being impact assessments.

Please describe the practical application of an enhanced well-being impact assessment framework for a/is that could be employed to address ethical and rights-based normative principles in AI. This process could enable a human-centered algorithmically-supported approach to the understanding of the impacts of AI systems.

38.2.10

Liu, HY., Zawieska, K. (2020) From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence. *Ethics Inf Technol* 22, 321–333 (2020). <https://doi.org/10.1007/s10676-017-9443-3>

As the aim of the responsible robotics initiative is to ensure that responsible practices are inculcated within each stage of design, development and use, this impetus is undergirded by the alignment of ethical and legal considerations towards socially beneficial ends. While every effort should be expended to ensure that issues of responsibility are addressed at each stage of technological progression, irresponsibility (meaning a lack of responsibility) is inherent within the nature of robotics technologies from a theoretical perspective that threatens to thwart the endeavor. This is because the concept of responsibility, despite being treated as such, is not monolithic: rather this seemingly unified concept consists of converging and confluent concepts that shape the idea of what we colloquially call responsibility. From a different perspective, robotics will be simultaneously responsible and irresponsible depending on the particular concept of responsibility that is foregrounded: an observation that cuts against the grain of the drive towards responsible robotics.

This problem is further compounded by responsible design and development as contrasted to responsible use. From a different perspective, the difficulty in defining the concept of responsibility in robotics is because human responsibility is the main frame of reference. Robotic systems are increasingly expected to achieve the human-level performance, including the capacities associated with responsibility and other criteria which are necessary to act responsibly. This subsists within a larger phenomenon where the difference between humans and non-humans, be it animals or artificial systems, appears to be increasingly blurred, thereby disrupting orthodox understandings of responsibility.

This paper seeks to supplement the responsible robotics impulse by proposing a complementary set of human rights directed specifically against the harms arising from robotic and artificial intelligence (AI) technologies. The relationship between responsibilities of the agent and the rights of the patient suggest that a rights regime is the other side of responsibility coin. The major distinction of this approach is to invert the power relationship: while human agents are perceived to control robotic patients, the prospect for this to become reversed is beginning.

As robotic technologies become ever more sophisticated, and even genuinely complex, asserting human rights directly against robotic harms become increasingly important. Such an approach includes not only developing human rights that 'protect' humans (in a negative, defensive, sense) but also 'strengthen' people against the challenges introduced by robotics and AI (in a positive, empowering, manner) [This distinction parallels Berlin's negative and positive concepts of liberty (Berlin, in *Liberty*, Oxford University Press, Oxford, 2002)], by emphasizing the social and reflective character of the notion of humanness as well as the difference between the human and nonhuman. This will allow using the human frame of reference as constitutive of, rather than only subject to, the robotic and technologies, where it is human and not technology characteristics that shape the human rights framework in the first place.

Try describe proposing a complementary set of human rights directed specifically against the harms arising from robotic and artificial intelligence (AI) technologies.

38.3 Additional resources for further Study III.

38.3.1

Silvia Milano, Mariarosaria Taddeo & Luciano Floridi (2020) Ethical aspects of multi-stakeholder recommendation systems. Pages 35-45, <https://doi.org/10.1080/01972243.2020.18326368>

In this article we analyze the ethical aspects of multistakeholder recommendation systems (RSs). Following the most common approach in the literature, we assume a consequentialist framework to introduce the main concepts of multistakeholder recommendation. We then consider three research questions: Who are the stakeholders in a RS?

How are their interests taken into account when formulating a recommendation? And what is the scientific paradigm underlying RSs? Our main finding is that multistakeholder RSs (MRSs) are designed and theorized, methodologically, according to neoclassical welfare economics. We consider and reply to some methodological objections to MRSs on this basis, concluding that the multistakeholder approach offers the resources to understand the normative social dimension of RSs.

What do you think on multistakeholder approach to AI? Does it have a significant advantages over the more traditional user-centred approach?

 38.3.2

Bertrand Braunschweig (Ed.), Malik Ghallab (Ed.) Reflections on Artificial Intelligence for Humanity (Lecture Notes in Computer Science Book 12600) 1st ed. 2021 Edition, Kindle Edition by Bertrand Braunschweig (Editor), Malik Ghallab (Editor) Format: Kindle Edition. Part of: Lecture Notes in Computer Science (129 books) <https://www.springerprofessional.de/reflections-on-artificial-intelligence-for-humanity/18838634>

We already observe the positive effects of AI in almost every field, and foresee its potential to help address our sustainable development goals and the urgent challenges for the preservation of the environment. We also perceive that the risks related to the safety, security, confidentiality, and fairness of AI systems, the threats to free will of possibly manipulative systems, as well as the impact of AI on the economy, employment, human rights, equality, diversity, inclusion, and social cohesion need to be better assessed.

The development and use of AI must be guided by principles of social cohesion, environmental sustainability, resource sharing, and inclusion. It has to integrate human rights, and social, cultural, and ethical values of democracy. It requires continued education and training as well as continual assessment of its effects through social deliberation.

After reading the book the “reflections on AI for humanity” please answer to the question:

- 1. How can we ensure the security requirements of critical applications and the safety and confidentiality of data communication and processing? What techniques and regulations for the validation, certification, and audit of AI tools are needed to develop confidence in AI? How can we identify and overcome biases in algorithms? How do we design systems that respect essential human values, ensuring moral equality and inclusion?**
- 2. What kinds of governance mechanisms are needed for personal data, metadata, and aggregated data at various levels?**
- 3. What are the effects of AI and automation on the transformation and social division of labor? What are the impacts on economic structures? What proactive and accommodation measures will be required?**
- 4. How will people benefit from decision support systems and personal digital assistants without the risk of manipulation? How do we design transparent and intelligible procedures and ensure that their functions reflect our values and criteria? How can we anticipate failure and restore human control over an AI system when it operates outside its intended scope?**
- 5. How can we devote a substantial part of our research and development resources to the major challenges of our time such as climate, environment, health, and education?**

38.3.3

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. European Journal of Education, 57, 632– 645. <https://doi.org/10.1111/ejed.12527>

In the Research described the augmentation perspective and the concept of hybrid intelligence are introduced to frame our thinking about AI in education. Molenaar, I. (2022).

„Education is a unique area for application of artificial intelligence (AI). In this article, the augmentation perspective and the concept of hybrid intelligence are introduced to frame our thinking about AI in education. The involvement of quadruple helix stakeholders (i.e., researchers, education professionals, entrepreneurs, and policymakers) is necessary to understand the opportunities and challenges of different educational use cases from an integrated point of view. To facilitate a meaningful dialogue, a common language about AI in education is needed.

This article outlines elements of such a common language. The *detect-diagnose-act* framework is used to describe the core functions of AI in education. The *six levels of automation* model is introduced to develop our thinking about the roles of AI, learners, and teachers in educational arrangements. In this model, the transition of control between teacher and technology is articulated at different levels and related to the augmentation perspective. Finally, the future of AI in education is discussed using self-regulated learning as an example. The proposed common language will help to support a coordinated development of an interdisciplinary dialogue between quadruple helix stakeholders to strengthen meaningful application of AI for learning and teaching.” Molenaar, I. (2022).

Please, describe the six levels of automation model develops our thinking about roles of AI, learners, and teachers in educational use cases.

38.3.4

Selwyn, N. (2022). The future of AI and education: Some cautionary notes. European Journal of Education, 57, 620– 631. <https://doi.org/10.1111/ejed.12532>

In light of fast-growing popular, political and professional discourses around AI in education, this article outlines five broad areas of contention that merit closer attention in future discussion and decision-making. These include: (1) taking care to focus on issues relating to 'actually existing' AI rather than the overselling of speculative AI technologies; (2) clearly foregrounding the limitations of AI in terms of modelling social contexts, and simulating human intelligence, reckoning, autonomy and emotions; (3) foregrounding the social harms associated with AI use; (4) acknowledging the value-driven nature of claims around AI; and (5) paying closer attention to the environmental and ecological sustainability of continued AI development and implementation.

Thus, in contrast to popular notions of AI as a neutral tool, the argument is made for engaging with the ongoing use of AI in education as a political action that has varying impacts on different groups of people in various educational contexts. Other researchers conclude that in contrast to popular notions of AI as a neutral tool, the argument is made for engaging with the ongoing use of AI in education as a political action that has varying impacts on different groups of people in various educational contexts. Selwyn, N. (2022). „In light of fast-growing popular, political and professional discourses around AI in education, this article outlines five broad areas of contention that merit closer attention in future discussion and decision-making.

These include: (1) taking care to focus on issues relating to 'actually existing' AI rather than the overselling of speculative AI technologies; (2) clearly foregrounding the limitations of AI in terms of modelling social contexts, and simulating human intelligence, reckoning, autonomy and emotions; (3) foregrounding the social harms associated with AI use; (4) acknowledging the value-driven nature of claims around AI; and (5) paying closer attention to the environmental and ecological sustainability of continued AI development and implementation. Thus, in contrast to popular notions of AI as a neutral tool, the argument is made for engaging with the ongoing use of AI in education as a political action that has varying impacts on different groups of people in various educational contexts.” Selwyn, N. (2022).

In light of fast-growing popular, political and professional discourses around ai in education, and based on reading article please outlines five broad areas of contention that merit closer attention in future discussion and decision-making.

Study Assignments Social and Ethical Aspects of AI

Chapter 39

39.1 Serie 1

39.1.1

Study Assignment 1: AI for Peace

At the level of policy making you may find more detailed non-technical guidance that has been produced by the '[AI for Peace](#)' group: Please check its reports as below and choose two of them that appeal most to you, in terms of academic interest, career options or societal commitment.

1. White Paper on Artificial Intelligence: a European approach to excellence and trust, European Commission, [February 2020](#)
2. Principled Artificial Intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI, Berkman Klein Center, [January 2020](#)
3. The role of artificial intelligence in achieving the Sustainable Development Goals, [2020](#)
4. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute, [2019](#)
5. UN Activities on Artificial Intelligence (AI), ITU, [2019](#)
6. Measuring What Matters in the Era of Global Warming and the Age of Algorithmic Promises, IEEE, [2019](#)
7. Artificial Intelligence and International Affairs: Disruption Anticipated, CHATHAM HOUSE, [2018](#)
8. The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation, [2017](#)

In the light of the recent (2022 and 2023) geo-strategic conflicts and potential conflicts; What did you learn about the instrumental impact of AI in conflict arousal, weaponed conflicts, its role in conflict mitigation and peace-keeping measures?

39.1.2

Study Assignment 2: Harnessing Collective Intelligence

Amazon Mechanical Turk (MTurk) <https://www.mturk.com> is a crowdsourcing marketplace that makes it easier for individuals and businesses to outsource their processes and jobs to a distributed workforce who can perform these tasks virtually. This could include anything from conducting simple data validation and research to more subjective tasks like survey participation, content moderation, and more. MTurk enables companies to harness the collective intelligence, skills, and insights from a global workforce to streamline business processes, augment data collection and analysis, and accelerate machine learning development.

While technology continues to improve, there are still many things that human beings can do much more effectively than computers, such as moderating content,

performing data deduplication, or research. Traditionally, tasks like this have been accomplished by hiring a large temporary workforce, which is time consuming, expensive and difficult to scale, or have gone undone. Crowdsourcing is a good way to break down a manual, time-consuming project into smaller, more manageable tasks to be completed by distributed workers over the Internet (also known as 'microtasks').

Imagine that you want to outsource a machine-learning task to an external enterprise. Explore and try out the building, managing, and evaluation of how Machine Learning workflows in MTurk can work for you. To what extent is the underlying assertion correct for you?

" MTurk can be a great way to minimize the costs and time required for each stage of ML development. It is easy to collect and annotate the massive amounts of data required for training machine learning (ML) models with MTurk. Building an efficient machine learning model also requires continuous iterations and corrections. Another usage of MTurk for ML development is human-in-the-loop (HITL), where human feedback is used to help validate and retrain your model. An example is drawing bounding boxes to build high-quality datasets for computer vision models, where the task might be too ambiguous for a purely mechanical solution and too vast for even a large team of human experts."

After your try-out of MTurk, ask yourself "What is in for me? For my study strategies? For building my professional network? For raising my own company? Etc. Dare to be critical and provide suggestions for a next-generation version of MTurk. (2 Pages max).

39.1.3

Study Assignment 3: The realism of a synthetic person

According to [Wikipedia](#):

..... A generative adversarial network (GAN) is a class of machine learning frameworks designed by Ian Goodfellow and his colleagues in June 2014. Two neural networks contest with each other in the form of a zero-sum game, where one agent's gain is another agent's loss. Given a training set, this technique learns to generate new data with the same statistics as the training set. For example, a GAN trained on photographs can generate new photographs that look at least superficially authentic to human observers, having many realistic characteristics. Though originally proposed as a form of generative model for unsupervised learning, GANs have also proved useful for semi-supervised learning, fully supervised learning, and reinforcement learning.



Consult <https://thispersondoesnotexist.com> and check and describe its potential benefits and risks.

39.1.4

Study Assignment 4: Bridging Humans and Machines

Affectiva's emotion database has now grown to nearly 6 million faces analysed in 75 countries. To be precise, we have now gathered 5,313,751 face videos, for a total of 38,944 hours of data, representing nearly 2 billion facial frames analysed.

Affectiva Automotive AI

- The leading In-Cabin Sensing solution that understands what is happening inside of a vehicle. It measures in real time, the state of the cabin, and that of the driver and occupants in it, to improve road safety and unlock personalized and more comfortable mobility experiences.

Affectiva Media Analytics

- It helps businesses understand how their customers and consumers feel when they can't or won't say so themselves. By measuring unfiltered and unbiased responses, businesses can act to improve customer experience and marketing campaigns.

In Lab Biometric Solution

- Gain a holistic view of human behaviour in your research with our partner iMotions' solution. Combined, this research platform seamlessly integrates emotion recognition technology and biometric sensors in one place.

" .. Affectiva's technology has validated exclusive measures to give confidence in market performance. Giving you clear guidance on targeting media groups, alongside executional challenges of improving story flow, creating cut-downs, testing voice overs and understanding the emotional role of your brand. As viewers watch your stimulus, we measure their moment-by-moment facial expressions of emotion. All

your research panellists need is internet connectivity and a standard web camera – it's simple, easy and highly accurate. The results are aggregated and displayed in an easy to use dashboard. Our norms help benchmark how your ads perform compared to those of your competitors – by geography, product category, media length and on repeat view. Unparalleled in the industry, these norms are built on the world's largest emotion database."

If you go to [Affectiva Media Analytics for Ad Testing](#) you will find the rationale for allowing AI to sense customers' eye response and measure the appeal of advertorials. Affectiva pretends to have the world's largest emotion database. How would you evaluate the Dashboard of Affectiva in terms of effectiveness, its consequences for guaranteeing privacy, and Kantar's Sustainability Sector Index, Link advertising database, and Affectiva's Emotion AI. Do you experience Affectiva's claim that these tools work in synergy to provide cohesive insights into consumers' responses to sustainability and a deep dive into top and bottom performing ads that featured social and environmental issues?

39.1.5

Study Assignment 5: Automatic Analysis of EEGs

This Technology Report announces: "Brain monitoring combined with automatic analysis of EEGs provides a clinical decision support tool that can reduce time to diagnosis and assist clinicians in real-time monitoring applications (e.g., neurological intensive care units). Clinicians have indicated that a sensitivity of 95% with specificity below 5% was the minimum requirement for clinical acceptance. In this study, a high-performance automated EEG analysis system based on principles of machine learning and big data is proposed."

In the full article that [you can download](#), the conclusion paragraph on Page 12 phrases:

".. Our hybrid HMM/deep learning system delivered a sensitivity above 90% while maintaining a specificity below 5%, making automated analysis a viable option for clinicians. This framework for automatic analysis of EEGs can be applied in other classification tasks such as seizure detection or abnormal detection. There are many straightforward extensions of this system that can include more powerful deep learning networks such as Long Short-Term Memory Networks or Convolutional Neural Networks. This is the subject of our ongoing research..."

If you take a critical stand to the architecture of the proposed HMM/deep learning system: A three-pass architecture for automatic interpretation of EEGs that integrates hidden Markov models for sequential decoding of EEG events with deep learning for decision-making based on temporal and spatial context. What alternative approach would you propose for a next generation for detecting and classifying EEG events? Hint: Consider that already since more than 20 years from thousands of patients' EEGs have been taken and accompanied with their medical progress/decline.

 39.1.6

Study Assignment 6: The Dueros Conversation

The Dueros announcements (<https://dueros.baidu.com/en/index.html>) mention

“ .. DuerOS is a conversational AI system developed by Baidu's Duer Business Uni. DuerOS is one of the leading applications of Baidu's world-class AI technology. Built on Baidu's large ecosystem of information and services, DuerOS delivers comprehensive solutions by allowing users to command and converse with their devices using natural language. As an open operating system, DuerOS also automatically learns at any given time to enable any machine to master human language. DuerOS intelligent chips have low production costs reduce power dissipation and are highly integrated. DuerOS is a versatile AI conversational system, and the intelligent chips can be integrated with any smart product including intelligent toys, Bluetooth speakers, small-home appliances, and more. It greatly lowers barriers for partners to use conversational AI system. DuerOS partners with various industries to incorporate artificial intelligence technology into people's everyday lives. With the advancement of AI technology, conversational systems like DuerOS can make it more convenient for people to obtain information and enjoy the benefits of technology.”

As credibility factor, the Duer enterprise asserts:

“ .. Baidu began investing in AI at a very early stage and has recruited top talent from around the world. It possesses unique advantages in the three key factors that drive today's AI revolution : algorithms, computing, and data. Baidu's algorithms are built on top of ultra-large-scale neural networks, trillions of parameters, and hundreds of billions of samples. Baidu's computing power stems from hundreds of thousands of servers and the largest high performance GPU cluster in China. As the world's largest Chinese search engine, Baidu has access to more than a trillion webpages, billions of search queries, images, video content, and positioning data. DuerOS synthesizes the best of Baidu technologies -- speech recognition, image recognition, natural language processing, user profile, and other advanced technical skills -- to create one of the most advanced conversational computing platforms available today...”

Looking into Duer's three key assets:

- **Algorithms:** Baidu's algorithms are built on top of ultra-large-scale neural networks, trillions of parameters, and hundreds of billions of samples.
- **Computing:** Baidu's computing power stems from hundreds of thousands of servers and the largest high performance GPU cluster in China.
- **And Data:** Baidu has access to more than a trillion webpages, billions of search queries, and tens of billions of images and video content, and positioning data.

Which of the three looks most realistic?

- **Speech Recognition:** Baidu's speech recognition technology has an accuracy rate of more than 97%.
- **Image Recognition:** Baidu's facial recognition system has an accuracy rate of more than 99.7%.
- **Natural Language Processing:** Baidu's AI has the ability to communicate with users via natural language and can understand users' intention with ease.

39.1.7

Study Assignment 7: [AI-powered K-12 Adaptive Instruction](#)

Squirrel Ai Learning is a K12 EdTech company specializing in intelligent adaptive education. The Squirrel Ai adaptive learning engine based on the advanced algorithm developed by an international team of scientists, designers and curriculum experts. Squirrel is the symbol for "agility, diligence and management." This aligns with the experience Squirrel Ai Learning provides for its students, to help them advance learning through the real-time adaptive system and cultivate good learning habits with practice. Like the AlphaGo simulated Go master, the AI system simulated human teacher giving the student a personalized learning plan and one-on-one tutoring, with much higher efficiency than traditional instructions. Squirrel Ai Learning offers the high-quality after-school courses in subjects such as Math, English, Physics, and Chemistry. Powered by its proprietary AI-driven adaptive engine and custom-built courseware, the Squirrel Ai Learning platform provides students with a supervised adaptive learning experience that has been proven to improve both student efficacy and engagement across its online learning platform and offline learning centers.

Squirrel Ai Learning pretends to enhance Primary and Secondary Education through artificial intelligence in China, based upon MIT-developed technologies: "The goal is to diagnose a student's gaps in understanding as precisely as possible" as formulated by Karen Hao. The crucial phrase in its further explanation on how Squirrel works it says: "Middle school math, for example, is broken into over 10,000 atomic elements, or "knowledge points," such as rational numbers, the properties of a triangle, and the Pythagorean theorem." What critical reflections can you offer against Squirrel's chosen architecture to start from "knowledge points"? What entities would you propose for building a next generation adaptive learning support system?

 39.1.8

Study Assignment 8: [Watson Assistant for Compliance?](#)

Watson Assistant: Build better virtual agents, powered by AI, and deliver consistent and intelligent customer care across all channels and touchpoints with conversational AI.

IBM's Watson can successfully guide end users through a complete action. It's now on the path to becoming the dynamic solution you envisioned when you first started the planning process. In our post on planning your assistant, we encouraged you to take a walk-run-fly approach. Your assistant is like a toddler that's just beginning to walk. It needs experience and guidance to learn how to run and, eventually, how to fly. Now that you've built your first action, the focus will shift to managing your assistant's lifecycle, which will consist of deploying and improving your assistant's content. Ensuring the long-term health of your assistant requires analysis, iteration, and reiteration. To empower you to achieve this flow, we've designed Watson Assistant with two separate environments:

- Draft, where you can continuously preview and edit your content
- Live, which contains the version of your content end users see when interacting with your assistant

You've already gotten a taste of how these environments interact while building your first assistant. Now, you'll manage your assistant by continuously building, deploying, and improving new versions. The navigation panel is organized along those lines:

- Build
- Actions
- Preview
- Deploy
- Publish
- Environments
- Improve
- Analyze

Go for instance to '[Student Loans](#)' and try out the 'Virtual Assistant'. What strengths and weaknesses do you experience? Where do you sense the AI underpinning of Watson's dialogue?

39.1.9

Study Assignment 9: [A.I. Emulates Human Behaviour](#)

When artificial intelligence systems start getting creative, they can create great things – and scary ones. Take, for instance, an AI program that let [web users compose music](#) along with a [virtual Johann Sebastian Bach](#) by entering notes into a program that generates Bach-like harmonies to match them. [Run by Google](#), the app [drew great praise](#) for being ground-breaking and fun to play with. It also attracted [criticism](#), and raised concerns about AI's dangers. My study of how [emerging technologies affect people's lives](#) has taught me that the problems go beyond the admittedly large concern about [whether algorithms](#) can really [create music](#) or art in general. Some complaints seemed small, but really weren't, like observations that Google's AI was [breaking basic rules](#) of music composition. In fact, efforts to have computers mimic the behaviour of actual people can be confusing and potentially harmful.

Google's program analyzed the notes in 306 of Bach's musical works, finding relationships between the melody and the notes that provided the harmony. Because Bach followed strict rules of composition, the program was effectively learning those rules, so it could apply them when users provided their own notes.

Go to the [virtual Bach composer](#) and describe

1. what harmonic sequences makes you think of the real Johan Sebastian Bach?
2. What will be needed further in order to make the virtual Bach music more excited?
3. Open the [Open AI Text Generator](#) and ask yourself the same questions

39.1.10

Study Assignment 10: [Introducing the First Serving Cobot](#)

Plato's role is not to replace waiters but to help and assist them in order to better their everyday life. Plato takes care of unpleasant tasks which allows waiters to have more time with clients.

Osteria

- "We quickly understood that Plato is not there to replace Man. But it acts more as a support to make our work easier and lighter." Julien, Waiter, Paris (France)

Howard Hotel

- "At breakfast time, the team is relieved by the service. As for the clients, there is curiosity, because Plato is not something that you see everywhere."
Baya, Manager, Paris (France)

Henriette

- "Plato allows the waitresses to be left in the dining room with the clients to be able to advise them, leaving to Plato the tasks without added value."
Guillaume, restaurant Manager, Beaucozé (France)

Download the PLATO virtual waiter via [the QR code](#) on your telephone, and

1. **check its demonstration of 'augmented reality'.**
2. **[Ask for a trial](#) and check what the remaining weaknesses are in comparison to a human waiter**
3. **Check [the humanity page](#) and ask yourself how virtual nurses and care workers for the elderly will behave in the near future**

39.2 Serie 2

39.2.1

Study Assignment 11: Robot Designed to Interact

Pepper is the world's first social humanoid robot able to recognize faces and basic human emotions. Pepper was optimized for human interaction and is able to engage with people through conversation and his touch screen. Pepper is available today for businesses and schools. Over 2,000 companies around the world have adopted Pepper as an assistant to welcome, inform and guide visitors in an innovative way.

- Standing 120cm tall, Pepper has no trouble in perceiving his environment and entering into a conversation when he sees a person.
- The touch screen on his chest displays content to highlight messages and support speech.
- His curvy design ensures danger-free use and a high level of acceptance by users.

1. [German United Robotics Group acquires SoftBank Robotics Europe in Paris](#)

Bochum / Paris / Tokyo, 12 April 2022

2. [Pepper x Marina Mall Abu Dhabi \(Robotic Gelato Bar\)](#)

In the current times when social distancing and contactless service have become the new normal, Pepper has joined hands with Eka and Tony robots at Abu Dhabi's first gelato bar entirely staffed

3. [100 Pepper cheerleading squad wins GUINNESS WORLD RECORDS™ certificate!](#)

Visit the three 'success stories' of Pepper and reflect upon the next questions

- What exactly makes anthropomorphic ('humanoid') robots so exciting for [children](#), [adults](#) and [the elderly](#)?
- Please, visit the page on [roboting for persons with the autism spectrum](#) and ask yourself what is crucial in AI-based robots for achieving at the aspects of: 1. Expressive and Communication Skills. 2. Cognitive and Emotional Skills. 3. Social and Interaction Skills, and 4. Skills for Acquiring Knowledge.
- Check how children tend to [learn with AI-driven robots](#); What keywords seem appropriate to describe its educational effects: Motivation, Curiosity, Self-Efficacy, Stimulating, Individualistic / Collaborative Learning, etc.

 39.2.2

Study Assignment 12: [Virtual, Augmented and Mixed Reality!](#)

Innovation with a Creative Touch: Virtual, Augmented and Mixed Reality. VRMonkey is a reference in the development of Virtual, Augmented and Mixed Reality experiences. We develop customized solutions so that our customers' games generate engagement and stand out from their competitors. We have a highly specialized team and projects of international prominence. As AI systems become more 3D spatial and immersive (as if you feel completely 'flooded' by the surrounding fictitious world), the question becomes 'isn't the real physical world enough for learning'? This is a good question as it makes us keen on what essentially is 'care', 'education', 'leisure time', 'working' etc.?

Visit and observe VRMonkey and experience the potential of immersion; What is your option on the contribution of 3D environments; Will it help to 'open your mind' for new aspects of life? Can it be used for changing attitudes, mentality and moral?

39.2.3

Study Assignment 13: [Science is Everywhere](#)



Great minds

Get to know some famous scientists from history



'Science is Everywhere' it incorporates experiments as created by artists and scientists. Its goal is to 'open door' to learners who would never have the opportunity to actually 'visit' these most famous places on earth and adjust their imagination to 'full reality'.

What educational scenarios (from direct instruction to exploratory and discovery learning) would you recommend in order to let learner fully benefit from the described perceptual assets?

39.2.4

Study Assignment 14: [One-to-One Teaching](#)

Closing the math attainment gap with one to one teaching is a dream as once in 'intelligent instructional science'. Math support wherever your school needs it. At Third Space Learning, we're on a mission to help teachers like you shape your pupils into confident, able mathematicians.

Whatever you and your school need to get there, we've got you covered. For Instance:

- DfE-approved tutoring programmes
- Free primary math resources
- Premium math resources and CPD

Please open “[Plug gaps and boost confidence across KS2](#)” and verify if the claimed benefits are realistic”

- “ Plug gaps and boost confidence across KS2. Personalized one to one math’s tutoring programmes designed and delivered by passionate math’s specialists..”
- “ Provide one to one tuition for your target pupils. Build engagement in a low stake learning environment. Boost confidence and familiarity with key content. Easily identify and plug gaps as far back as KS1. Develop reasoning skills and boost verbal fluency.....”
- “... An excellent KS2 math’s intervention based on assessment for learning. Pupils have personalized support with minimal impact on staff workload....”

 39.2.5

Study Assignment 15: [Ada Lovelace Institute](#)

At the Ada Lovelace Institute we use a range of research methodologies to explore sociotechnical questions relating to data and AI.



Biometrics

Exploring how the governance of biometric data and technologies can be strengthened to keep pace with new applications and ethical questions



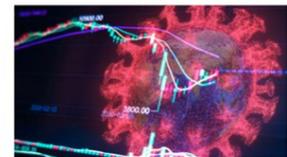
Public sector use of data and algorithms

Building evidence on how data-driven technologies are shaping public services, and supporting approaches to public-sector algorithm accountability



The future of regulation

Forecasting and shaping the role of regulation in mediating power over data and AI for the good of people and society



Health data and COVID-19 technologies

Exploring the impact of health technologies on people and society in the context of the COVID-19 pandemic

Ada Lovelace Institute About Our work Ada in Europe Blog News & events JUST AI 



Project

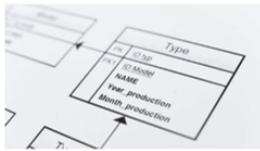
Algorithmic impact assessment in healthcare

A research partnership with NHS AI Lab exploring the potential for algorithmic impact assessments (AIAs) in an AI imaging case study

10 March 2021

Accountability mechanisms

AI bias ...



Project

Algorithmic accountability for the public sector

Research with AI Now and the Open Government Partnership to learn from the first wave of algorithmic accountability policy

24 March 2021

Accountability mechanisms

AI policy ...



Project

Independent legal review of the governance of biometric data in England and Wales

An independent legal review of the governance of biometric data, commissioned by the Ada Lovelace Institute and led by Matthew Ryder QC

Biometric data

Biometric technologies ...



Project

Public dialogue on the ethics of location data

A partnership with Traverse, the Geospatial Commission and Sciencewise to understand public perspectives on the responsible use of location data

18 March 2021

Data governance Location data

...

1. Open the section: '[Algorithmic impact assessment in healthcare](#)' and check its premise "...A research partnership with NHS AI Lab exploring the potential for algorithmic impact assessments (AIAs) in an AI imaging case study"
2. Open the section: '[Independent legal review of the governance of biometric data in England and Wales](#)' and its '[Policy Report](#)' and check its premise ".... It draws from Ada's public engagement research on attitudes towards biometric technologies, an independent legal review led by Matthew Ryder QC, and desk research to provide background on current developments in the realm of biometric technologies and their governance. It puts forward a set of ambitious policy recommendations, that are primarily for policymakers and will also be of interest to civil-society organisations and academics working in this contested area...."

39.2.6

Study Assignment 16: [Ethics-Governance by A. I.](#)

Launched in 2017, the Ethics and Governance of AI Initiative is a hybrid research effort and philanthropic fund that seeks to ensure that technologies of automation and machine learning are researched, developed, and deployed in a way which vindicate social values of fairness, human autonomy, and justice. The Initiative is a joint project of the [MIT Media Lab](#) and the [Harvard Berkman-Klein Centre for Internet and Society](#). It incubates a range of research, prototyping, and advocacy activities within these two anchor institutions and across the broader ecosystem of civil society.

Please provide the answers to the next three questions, to the best of your knowledge and 'educated guess':

- **AI and justice** - What legal and institutional structures should govern the adoption and maintenance of autonomy in public administration? How might approaches such as causal modelling rethink the role that autonomy has to play in areas such as criminal justice?
- **Information quality** - Can we measure the influence that machine learning and autonomous systems have on the public sphere? What do effective structures of governance and collaborative development look like between platforms and the public? Can we better ground discussions around policy responses to disinformation in empirical research?
- **Autonomy and interaction** - What are the moral and ethical intuitions that the public brings to bear in their interactions with autonomous systems? How might those intuitions be better integrated on a technical level into these systems? What role does design and interface - say, in autonomous vehicles - play in defining debates around interpretability and control?

 39.2.7

Study Assignment 17: [Social Implications of AI](#)

The AI Now Institute aims to produce interdisciplinary research and public engagement to help ensure that AI systems are accountable to the communities and contexts in which they're applied.

1. Their mission is to produce rigorous, interdisciplinary, and strategic research to inform public discourse around the social implications of AI.
2. Breaking the Lock: Accessing Public Records to Map Systems, Algorithms and Data
3. Water Justice and Technology Report
4. The Steep Cost of Capture
5. Democratize AI? How the proposed National AI Research Resource falls short
6. Algorithmic Accountability for the Public Sector Report
7. Smartphones, sensors and consumer habits reveal much about society. Too few people have a say in how these data are created and used.

Please open '[Breaking the Lock: Accessing Public Records to Map Systems, Algorithms and Data](#)' and question your knowledge in how far the underlying constraints can be met.

1. **Review the federal Freedom of Information Act (FOIA) or Your State's Open Records Law – If you are filing a federal FOIA request, the Centre for Constitutional Rights has many useful tips available as part of our FOIA Basics Guide for Activists available at <http://foiabasics.org>. While many state public records laws are very similar to the federal law, several vary drastically so it is important follow this guidance in context of your state law. Two helpful resources are the [Open Records Guide\(link is external\)](#) provided by the Reporters' Committee for Freedom of the Press and also [Muckrock \(link is external\)](#).**
2. **Compile Sources or References to the System – Try to tailor your request in accordance with known or speculative information about the system. Potential sources of references to systems may include recent news articles, public statements or press releases by public officials, agency or legislative budgets, agency or legislative public hearing notes or minutes, past releases of FOIA documents or relevant databases (e.g. [MuckRock Project Public Records Request Archive\(link is external\)](#)). For example, after the murder of George Floyd in 2020, New York state issued Executive Order 103, which ordered all law enforcement agencies to review their individual policies and procedures and draft a reform plan for the future. Many of these plans were made public and may provide important information for your records request. Compiling these reference sources in advance may make your request more specific and help if you have to appeal or challenge the response to your public records request. Try to review the most recent information as possible.**

3. **Draft Public Records Request** – You can draft your request using the [annotated sample FOIA](#) or [annotated sample](#) public records request as a template but make sure to use your own sources or reference documents to narrow and craft your request. In addition to being specific, you should try to include examples of what you are looking for so that it is as clear as possible to the public records officer making determinations about what is responsive to your request. You can provide cites or links to these examples in your request, such as providing a link to a news article, or you can attach them to your request as exhibits if they aren't too long.

39.2.8

Study Assignment 18: [Teach morality to machines?](#)

Today, it is difficult to imagine a technology that is as enthralling and terrifying as machine learning. While media coverage and research papers consistently tout the potential of machine learning to become the biggest driver of positive change in business and society, the lingering question on everyone's mind is: "[Well, what if it all goes terribly wrong?](#)"

For years, experts have warned against the unanticipated effects of general artificial intelligence (AI) on society. Ray Kurzweil [predicts](#) that by 2029 intelligent machines will be able to outsmart human beings. Stephen Hawking argues that "[once humans develop full AI, it will take off on its own and redesign itself at an ever-increasing rate](#)". Elon Musk warns that AI may constitute a "[fundamental risk to the existence of human civilization](#)". Alarmist views on the terrifying potential of general AI abound in the media.

More often than not, these dystopian prophecies have been met with calls for a more ethical implementation of AI systems; that somehow engineers should imbue autonomous systems with a [sense of ethics](#). According to some [AI experts](#), we can teach our future robot overlords to tell right from wrong, akin to a "[Good Samaritan AI](#)" that will always act justly on its own and help humans in distress.

Although this future is still decades away, today there is [much uncertainty](#) as to how, if at all, we will reach this level of general machine intelligence. But what is more crucial, at the moment, is that even the [narrow AI applications](#) that exist today require our urgent attention in the ways in which they are making moral decisions in practical [day-to-day situations](#). For example, this is relevant when algorithms make decisions about who gets access to loans or when self-driving cars have to [calculate the value of a human life](#) in hazardous traffic situations.

1. **The first question is: "What happen if everything goes terribly wrong?"** Could you anticipate to the worst-case scenarios and mention generic measures to prevent from AI systems running out of hand?
2. **The prediction by Ray Kurzweil that by 2029 intelligent machines will outperform humans comes closer quicker than we thought.** Could you provide some arguments that in 2029 this prediction will not be met?

3. What is meant by “Good Samaritan AI”?
4. Would you call the algorithms in Facebook, Twitter, and Google “Narrow AI Applications”? If not; how far are we from full-blown AI? Explain how the situation in social media may look in ten years from now?

39.2.9

Study Assignment 19: [A.I. on the World Economic Forum](#)

The World Economic Forum proves to be quite keen on its rubric ‘[Artificial Intelligence](#)’.

- Artificial Intelligence is rife with contradictions. It is a powerful tool that is also surprisingly limited in terms of its current capabilities.
- Harnessing Artificial Intelligence to Accelerate the Energy Transition ...
- AI has tremendous potential to accelerate and support the global energy transition, but several key barriers are ...
- Chatbots RESET. This project brings together developers, platforms, the medical community, civil society, academia, and healthcare regulators to design ...
- Global Future Council on the Future of Artificial Intelligence. Council mission and objectives. The emergence of stronger forms of artificial intelligence, ...
- Artificial intelligence (AI) is a key driver of the Fourth Industrial Revolution. Its effect can be seen in homes, businesses, schools and even public spaces, ...
- A toolkit designed to help companies develop trustworthy AI for children and youth and to help parents, guardians, children and youth ...
- How Artificial Intelligence Could Widen the Gap Between Rich and ...
- To prevent this growing divergence, policymakers in developing economies will need to take actions to raise productivity and improve skills ...
- The New Physics of Financial Services – How artificial intelligence is ...
- Building on the World Economic Forum's past work on disruptive innovation in financial services, this report provides a comprehensive ...
- A Blueprint for Equity and Inclusion in Artificial Intelligence | World ...
- In this report, the World Economic Forum Global Future Council on Artificial Intelligence for Humanity has identified gap areas and ...
- Over the past decade, artificial intelligence (AI) has emerged as the software engine that drives the Fourth Industrial Revolution, ...

Take a look at its main three AI strands:

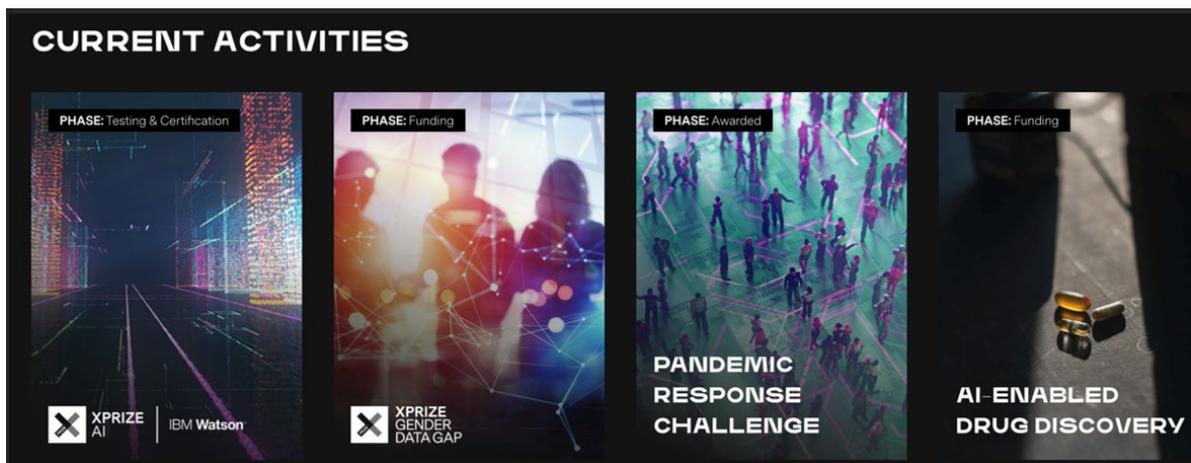
1. Discover [Artificial Intelligence for Children](#)
2. Participate [Sustainable Development Impact Meetings](#)
3. Take Action [Engage with the Centre for the Fourth Industrial Revolution](#)

Download the report attached to the strand of your main interest and summarize its key conclusions. How does it change your overall view on AI's societal impact?

 39.2.10

Study Assignment 20: [AI & Data for the Benefit of Humanity](#)

Initiated by XPRIZE along with visionary partners, AI & Data For Good Initiative is focused on enabling and cultivating actionable and tangible solutions to some of the world's most immediate challenges through the use of data sciences, machine learning and artificial intelligence. Built around a coalition of the most forward-thinking organizations globally, this initiative offers members an opportunity to evaluate and contribute to projects initiated by member organizations, assist with validation and visibility, promote access to standards groups, and leverage the AI for Good Global Summit and the UN platform for project visibility, impact, and scale. Through its live interactions and digital platforms, the AI & Data for Good platform provides a collaboration environment to launch projects and competitions, the outcomes of which directly address the UN's Sustainable Development Goals (SDGs) and aims to create an open participation and knowledge hub to use AI for the common good.



Please explore the four 'Current Activities' of XPRIZE and choose which of them has the highest new value for you. If you find it difficult to decide, please open the [page with the XPRIZE winners](#) and let yourself be inspired by their motivations and life missions. Make a half page report upon your reflections.

Notes and External Links

Chapter **40**

40.1 Resources

40.1.1

Subsidiary Links to AI Projects

1. <https://digitallibrary.io>
2. <https://www.changedyslexia.org> e.g. <http://www.voiceitt.com> , <https://www.nuance.com> , <https://otter.ai> and <https://kidsense.ai>
3. <https://blogs.microsoft.com/ai/ai-powered-captioning/>
4. <https://consumer.huawei.com/uk/campaign/storysign/>
5. An example of a robot developed for children on the autism spectrum is Kaspar (Dautenhahn et al., 2009). See, for example, Bughin et al., 2017; Frey and Osborne, 2017; Frontier Economics, 2018; Leopold et al., 2018; Madgavkar et al., 2019; and Manyika et al., 2017.
6. Manpower Group. 2016. Millennial Careers: 2020 Vision Facts, figures and practical advice from workforce experts. Available at <https://www.manpowergroup.com/wps/wcm/connect/660ebf65-144c-489e-975c-9f838294c237/>
7. Courses designed to enable citizens to become familiar with how AI works can be found at <https://www.elementsofai.com> , <https://okai.brown.edu> and <http://ai-4-all.org> .
8. Resources designed to help teachers introduce their students to AI can be found at <http://teachingaifork12.org> and <https://github.com/touretzkyds/ai4k12/wiki>
9. <http://www.gettingsmart.com/2018/07/coming-this-fall-tomontour-school-district-americas-first-public-school-ai-program>
10. <https://www.teensinai.com>
11. <https://www.skillsfuture.gov.sg/>
12. <https://microcompetencies.com>
13. <https://github.com/touretzkyds/ai4k12/wiki>
14. <http://teachingaifork12.org>
15. <https://www.elementsofai.com>
16. <https://okai.brown.edu>
17. <http://ai-4-all.org>
18. <https://www.oecd.ai/dashboards>
19. <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>

 40.1.2

Summary of what you are Supposed to Remember

1. What technical aspects have been listed under AI traditionally?
2. What are the major benefits for having this course module studied carefully?
3. What were the first three domains that were encoded into early AI systems? Which two games were equipped with AI and what fundamental understanding did they bring?
4. What wanted John Searle to clarify with his 'Chinese Room' metaphor? What is the meaning of 'keeping human operators in the loop?' Can you repeat the three main elements of 'situational awareness'?
5. What is essential for the transition Society 4.0 to 5.0? And for the transition Industry 3.0 to 4.0?
6. What is a 'cobot'? What 'levels of control' can we distinguish in roboting?
7. What side effects can we expect if full priority to technological autonomy is allowed?
8. Remember: Which AI service platforms can you mention? Which large AI-dependent enterprise is missing remarkably?
9. Which three major benefits from AI have been listed in the video by Peter Klagyivik? Which two controversial societal aspects of AI have been mentioned?
10. What is the role of intuition in innovative idea generation? How can AI help in it?
11. What is the cascade in machine intelligence? Why is 'unsupervised machine learning' so important?
12. Which of the seven technologies on the ladder from basic to advanced AI technologies can you mention 'by heart'?
13. Which intellectual tasks is AI relatively good in, and which prove to be much more difficult than we thought before?
14. What was given as the goal for Darpa's PRIDE system? What are its easier and more complex elements?
15. What percentage of the jobs have been estimated to disappear because of AI?
16. What is meant by dataveillance? and Ethics washing?
17. At what points is AI hailed and feared?
18. When was the term AI used for the first time? What is the typical trust in AI? Is it less or more than the average trust that customers have in their providers?
19. Explain the difference in accountability and empathy. Which one is easier to achieve by AI?
20. What proportion of respondents fear to become slave of the AI application? How many percent fears to lose his/her job to AI?
21. What is the biggest source of error in AI performance? Why would it be so difficult for AI to fully interact as a human person?
22. Which weaknesses are inherent to the current generation of AI systems?
23. What is crucial in a human – AI dialogue?

24. What is the main difference between the human existential rights and Asimov's 'Laws for a Robot'?
25. What is the goal of this course module? How does Hawkins conceive the end of the human race?
26. What can AI offer to cash-flow management?
27. What recent factor increased optimism on AI's future?
28. What is the role of ethicists in AI teams?
29. What is according to Accenture crucial in 'technology transformation'?
30. What data sees Payt as disputable data in an AI-oriented customer database?
31. What are the three levels of autonomy in machine learning?
32. What are spurious factors? Can you come up with your own example of a spurious factor and how this can damage scientific conclusions?
33. Do you agree that Google maps frees the driver from cognitive load? What other AI-driven applications for the consumer can you mention? Do you see negative side effects for the user in the long run?
34. What is 'learning for AI machines' in the most optimal sense?
35. What in the context of Singularity is a 'runway reaction'? Will it finally impair human intellect, or will it shift human thinking towards spiritual concerns?
36. What is the relation between AI and Science Fiction? Is AI more than mathematics and logical reasoning?
37. Starting from the premise of 'Strong AI', should we limit AI to the ultimate tasks that human experts can solve?
38. What is meant by 'Artificial Life' and what could it add the current AI systems?
39. What means Jürgen Schmidhuber as he postulates the 'formal theory of creativeness'? What would be the equivalent for 'creative AI'?
40. Which three major variables play a crucial role in the arousal of 'curiosity' due to Berlyne?
41. What is the major role of AI in designing Video Games?
42. What is the role of 'Explainable AI' (XAI), and how has this challenge been approached so far?
43. The Eradication scenario is a dramatic version of the AI take-over. Which one of the other four is the most influential one? Please argue why.
44. Neural Correlates of Consciousness or NCC has been postulated to explain how human awareness evolves. However, NCC heavily relies on a wide spectrum of physiological bodily processes. Would it be a good idea to equip AI consciousness with a complete emulation of the human body: respiration, heartbeat, hormones etc.? Is AI consciousness an obligatory ingredient before AI ethics can become achieved?
45. The 1943 choice to limit Neuronal computing to binary logic has been quite decisive for AI control structures since then. Quantum Computing (QC) offers a new dimension. Can you imagine what it QC may imply for AI finally?
46. Can you recollect why the metaphor of ant colony triggers the imagination of AI by parallel processing?
47. What could AI offer to protect nature? Less travelling? Less? Less spoiled energy? Less wars? Please formulate your thoughts based upon the prior video and Wikipedia fragment.

48. “ .. It would be a mistake, however, to suppose that cognition consists simply of building maximally accurate representations of input information” Does this mean that anthropomorphic robots are no longer needed?

40.1.3

Summary of thinking questions

1. Check the news media for recent events around A.I. Is it complementary to your earlier view on the essence of Artificial Intelligence? Feel welcome to share your findings with your peer students.
2. In this course module there is ample awareness on how citizens may become victim to AI mechanisms and will soon lose part of their privacy and autonomy. Have you thought about your own data in the broader context of AI like social media, cameras on the streets and squares? If so, please make a note in your agenda so that after few weeks you can compare the difference.
3. Complacency is often mentioned as basic attitude that makes citizens vulnerable: Policy makers, politicians and employers are then tempted to take a paternalistic stand and defend their tutelage by saying that citizens ‘need guidance’ in order not to become victim of naivety and backwardness. What is your opinion about how to make the right balance between tutelage and staying in a complacent ignorance?
4. If you look to Figure 1 above; What role do you estimate Education plays in the cascade from Society 4.0 to Society 5.0 and Industry 3.0 to Industry 4.0? If you have no clue, feel happy to consult the web and please watch the next video.
5. Each upward transition in Industry brings the need to upgrade workers’ qualifications. We have seen that craftsmanship in the transition Industry 3.0 to 4.0 has urged workers to incorporate ‘computational’ and ‘algorithmic’ thinking instead of having a sharp eye for size and material responses. What typical new demands for industrial workers do you estimate to arrive after the transition from Industry 4.0 to 5.0?
6. Given that Society 5.0 brings citizens in the condition of ‘Smart Cities’ with abundant sensors and AI for optimizing traffic streaming and reasoning about safety, environmental health etc. What transition do you estimate education to go through in terms of curricular content and instructional methods?
7. Try out each of the seven given AI platforms; What are its first glance strengths and weaknesses? What human task would you favour to support with AI? How would you divide its human- versus the machine component?
8. After having tried out Open AI and having read its Charter; What is the difference between consulting Google, Quora or Watson and OpenAI in its approach and achievement.
9. Thinking about AI-driven weapons and robots like drones. In case it hurts innocent civilians, who will be responsible? The builder of the drone, AI programmer, the person who launches the drone, or?

10. If you extrapolate the entering of grandparents in youngsters' life to the current situation, you may easily scan the Middle Ages, where monasteries and minstrels spread intellect and awareness. You might also easily pass through the era of Enlightenment, where the book printing press and formal education became necessary for spreading the fast scientific harvests. The arrival of early AI made us aware that human expertise could become consolidated in data systems and even in deductive reasoning machines. Could you try and extrapolate what catalytic role AI systems may have in the second half of the 21ST century? Think freely in terms of 1. supplanting the current model of transfer-oriented education? 2. allowing students to use computers in holistic / meaningful rather than rote memorization in school examinations? 3. Make learners keen on collaborative- rather than soloistic task performance? etc.
11. The plead against monotonous reasoning may also explain why teams of experts find even better solutions when one or two novices join the meetings and pose 'stupid' questions. Can you explain why?
12. In the paragraph above, the term machine 'learning' is questioned. What would be your criterion for calling AI as 'learning' process?
13. The last row in the Table 3 above lists 'Artificial Creativity'. The table suggests that this is the most demanding AI activity. Do you agree? Secondly: would you AI-driven random art label as 'Artificial Creativity' as well?
14. 'Manoeuvring through a garden' is labelled by Moravec as not trivial at all. For a 2-year youngster this task is estimated as relatively easy. What do you see as 'algorithmically' complex in 'Manoeuvring through a garden'?
15. The intelligence of switching on/off the heating based on actual temperature is seen as relatively easy. What factors should be considered if the criterion is that persons in that room should really feel a 'comfortable temperature'?
16. One may assume that AI supplants the high-paid jobs earlier than the low-paid jobs. Do you see exceptions to this assumption?
17. Questions:
18. What criteria should be considered in defining and continuously updating the ethical boundaries of the collection and use of learners' data?
19. How might schools, students, and teachers opt out from, or challenge, their representation in large datasets?
20. What are the ethical implications of not being able to easily interrogate how AI makes decisions (using multi-level neural networks)?
21. What are the ethical obligations of private organizations (product developers) and public authorities (schools and universities involved in research)?
22. How does the transient nature of students' interests and emotions as well as the complexity of the learning process impact on the interpretation of data and ethics of AI applied in educational contexts?
23. What pedagogical approaches are ethically warranted?
24. What ethical dilemmas do you see when AI would be included in decision taken by the city council?
25. AI is often supposed to fail because of its lack of empathy. Would it be solved if AI programmers would be trained to think emphatically?

26. The dimension accountability \Leftrightarrow empathy needs to be clarified even further. Please argue why and how human 'empathy' may have developed in evolution. Will the coming years happen the same with 'empathy' in AI systems?
27. It seems that the predicate 'intelligent' jobs has at least two aspects: 1. The complexity in reasoning and thus the needed amount of computation. And 2. The amount of overall empathy and human understanding like we see as decisive for good teachers, nurses, hairdressers, etc. Could you please argue if after the arrival of AI, the empathy-job will become better-paid?
28. The paragraph above argues that it is not a good idea to let AI emulate human tasks but make them complementary to the way human approach certain tasks. For instance, a shopkeeper might look to customers' faces when they look to the price tags on the shelves. For the computer, it is quite tedious to derive emotions from human faces. The shopkeeper might better use AI for finding patterns in customers' spending behaviour and subsequently find the optimal price per certain product. Could you think of another example that illustrates the trade-off between human- and artificial intelligence in business?
29. What examples can you remember of companies attempting the customer to understand how and why AI is needed? If you have no clue; please think of sectors like healthcare, education, car maintenance, insurance companies, etc.
30. Some AI-penetrated tasks are easily adopted, as the 'victim' has no choice. For instance, if you compile a CV and know from the beginning that an AI agent will sort the candidates. What is your opinion on slowly integrating AI in the citizen-government transactions? What is your recommendation for local and national authorities in this?
31. 'Critical Thinking' is mostly seen as attempt to fuller understand why phenomena and human conventions are as they are. If it comes to the regulation of AI, what analogy to human norms would you see as most natural?
32. If AI is going to rid us from spreadsheets; what will rid us from AI?
33. 'Cash Flow' sounds like a rather rigid computational job. One of its five elements is obviously much more human than algorithmic. Please describe how AI would perform on this aspect?
34. The arrival of web-based transactions like in the banking sector has made it easy to insert AI in banking. Still 'sitting around the table' is essential for creating trust. What situation do you remember where man-computer interaction was even more satisfactory than the human partnership?
35. Biased data are a notorious source for failing AI. If it comes to 'big data'; what problem is often met?
36. A strong example of AI making a contribution to governance is the consequential analysis in law-making: What exceptions need to be defined before a new law can fully become rolled out. Can you find another application where AI seems to work out well?
37. Why is 'unsupervised' learning that essential in the future of advance AI?
38. In order to prevent from 'profiling', citizens on the basis of race, social status or sexual orientation, it seems crucial not to have these variables accessible

- in investigation policies. From the other side, efficiency in Law enforcement is opportune as well. What rational would you apply for finding an acceptable balance between these two?
39. Vehicles so far help us to migrate persons and goods. The issue of smart mobility can be seen as just a symptom of modern societies to save time. As you consult the latest news on Metaverse, it may become clear that virtual presence will take over large parts of physical displacements. What will become the scenario for being in a self-driving car; Is it likely that the 'dismissed driver' stays virtually present with his/her origin? Or: is it more likely that the opposite is going to happen: Stay physically at your own location, and travel virtually around to place you want to visit?
 40. Neural networks have proven the power of 'learning by many examples'. Human learning also uses this 'informal learning'; not a guided process otherwise than social attention, curiosity and the will to survive and care for offspring. Can you think of more examples of neural learning than distinguishing dogs and cats?
 41. The premise at the start of the paragraph above is that AI will help to improve itself. We already know for two decades the software that generates the layout of PCBs (printed circuit boards) and electronic traces on silicon chips. However, more recently, the AI 'Alpha Go' succeeded to discover the 'rules of thumb' for winning the enormous complex middle stage in the Go game. It is a vivid exemplar that AI is feeding itself so that we might arrive earlier at the stage of singularity. Please check the coming week which other specimens of self-propelling AI are being announced via the media.
 42. The Wikipedia article on the origin of Artificial Intelligence mentions the emergence and decline of believing that computers could finally reach the level of human thinking. Looking back to the "AI winter" in the late seventies may make us more critical on the term 'Artificial Intelligence'. Seen the current industrial eagerness for machine learning, deep learning, data analytics and big data, we might better choose a more adequate term for AI. What would be a better term, in order to avoid common fear towards threatening singularity?
 43. Artificial General Intelligence (AGI) can be seen as the more ambitious attempt to let computer software reach the level of consciousness and even the level of 'wisdom'. What are your considerations to believe this high ambition; is it realistic; or is it simply beyond our cognitive capacity to think about it?
 44. Vladimir Red'ko poses the preference not to try and emulate the virtues of natural evolution like the human brains, but better try to study "processes-as-they-could-be". In other words, to excavate the mechanisms underlying natural evolution ('wet' intelligence) so that we are less obsessed (and thus 'hampered') by current silicon and software limitations. What alternative approaches come to your mind for actively undertaking the suggestion posed by Vladimir Red'ko? In case you have no idea, please consider notions from other literature like: 'self-organisation', 'quantum computing' and 'wisdom of the crowd'
 45. If you explore Schmidhuber's theory on rudimentary artificial science, what contribution to future AI would you see in creative art? In other words: What

contribution to current AI was given by Science Fiction in the 19th and mid 20th century (visionary writers like Jules Verne, H. G. Wells and Edgar Allan Poe)?

46. Berlyne's postulate of 'collative variables' brought him to the construct of 'conflict'. Does it explain that travelling raises more new ideas than sitting behind your desk? Talking about curiosity towards AI; What aspects of future AI attracts your imagination most?
47. Games encourage humans to think about the future as they pose fiction as temporary reality first of all. The term 'Virtual Reality' as 'fictitious world' slowly becomes accepted as the default 3D space that overcomes physical alienation as we experienced during video conferencing during the COVID-19 era. What means virtual reality for you; is it just a corollary of gaming and thus a way of escaping from full reality; or is it already an innate need for current teenagers?
48. Supervised machine learning allows the expert to stepwise regulate the amount of variables and dimensions to be taken into account in order to still 'understand' a certain AI conclusion. If you think of your own (human) personal learning in the past, you might need to accept that parts of your academic learning were not always 'transparent': you were not always able to understand the building blocks that you use for 'seeing the larger picture'. For instance, during COVID-19 it was not easy to understand how viruses propagated and how immunity developed. Can you go back to your own understanding of virology and confess what pervasive misconceptions you needed to give up finally?
49. Stephen Hawking, based upon a cosmic humility, claimed that there is no physical law precluding particles from being organized in ways that perform even more advanced computations than the arrangements of particles in human brains. What is your opinion on our progress to 'connect human minds' via 'electrodes' and 'future social media'? What options are there for future AI to play a stimulating role in this?
50. Which alternatives look plausible? Consciousness in AI seems to be far away, as it implies, a sense of ego, self-reflection, a degree of responsibility, the right for self-defense
51. Turing defined the needed tools for mathematics: paper, pencil, a working set of eyes and eraser. One could see a one-to-one isomorphy to the Neumann architecture where the processing unit reads a linear sequence of data and operations that could even make the reading on addresses forward and backward. The later programming paradigms attempted to stay at the level of 'rules' and 'facts' so that procedural prescriptions became minimal. If you have some experience with declarative languages like LISP or PROLOG, please provide an example that shows that procedural awareness is still needed.
52. Algorithmic programming allowed to iterate quite fast along computations, even entire matrix operations could be anticipated within one keystroke by the programmer, for instance in the IBM language APL. Later efforts evolved to base programming upon functional-, object-oriented- and aspect-oriented languages. If you monitor the trend in programming styles, do you see an indication that we get closer and closer to the needs in AI applications?

53. The mind-body dualism as the basis for mentalistic AI easily leads to ignore situational and bodily factors. Will AI perform better if these factors are included in the computational model?
54. Environmental-, diverse-, emancipatory-, democratic- and transparent are all buzz words in Western Society the last two decades. What do these trends imply for next generations AI?
55. When we accept bodily sensation and emotions as key factor in human cognition, we arrive closely to the role of 'intuition' in problem solving tasks. For instance 'estimating' the size of large areas may easily become blurred as its magnitude stretch far beyond what we can normally see, hear and feel. Can you give an example of how humans typically fail when estimating how long it takes light to travel to the moon?
56. When we accept bodily sensation and emotions as key factor in human decisions like in moral condemnations; how would AI manage to assist in jurisdiction?
57. If you go along all mentioned AI application domains above; which of them would you favour to work on in further study projects and in your later professional career?
58. Check the meaning of 'Cloud Foundry Service Platforms' and ask yourself how AI has the options to care and purify vital data?
59. The notion of collaborative robots have been coined. The human bodily interactions in for instance dancing might be captured for later artistic AI. What future applications can you predict?
60. Question and for Remembering. geographically distributed assets have been mentioned in the context of a bottom-up approach. How do you imagine its functioning for spare part inventory planning?
61. Knowledge discovery has been highlighted as we mentioned Alpha-Go. What message from the successes in Alpha-Go can go to AI-based Health Care?
62. The Wikipedia article above asserts that there are four major challenges in realizing industrial AI: data-driven, its speed, fidelity, and interpretability. The neural networking approach would complicate the criterion of interpretability. Could you describe in what sense interpretability will easily show its urgency in optimizing industrial processes; in particular at the aspect of fault finding.
63. Recently (mid 2022) the geopolitical forces tend to fuel the alertness to China's strategic need for AI in mastering nano-meter granularity of chip design. AI and micro-chip seem mutually dependent. How would cognitive representations like cortical interconnectivity decide upon the next generation of computers for self-driving cars for instance? What ethical constraints would you see for excluding China from this techno-race?
64. Scouting talents for the labour market is often seen as a case of matching CVs with job demands. A more fundamental question is if head-hunters should stay at the level of match making. Scouting and finding unheard talents will keep new colleagues at the level of the existing ones. Steve Jobs complained ironically: "Apple scouts the best young talents world-wide and pays them top salaries. Subsequently we tell them what to do ...". Please reflect on this dilemma: What would be the best ethical way to attract young talents without narrowing the enterprise's scope?

65. As AI research already reached the delicate level of weapon industry; can we still allow AI architectures to arrive in education and scientific publications?
66. Seen the evolutionary nature of human intellect; how can further AI development learn from animal to human cerebral development?
67. Turing's "polite convention": If a machine behaves as intelligently as a human being, then it is as intelligent as a human being. What abstraction legitimates Turing's thesis? In other words: what flaw can be seen in Turing's argumentation?
68. Final Questions:
69. Most technological inventions in the last five centuries have triggered the question if not too close the human soul would be touched:
70. The printing press, that would bypass human handwriting and spread the word to those who were never involved in the bible before. Initially, the reprinted books caught Suspicion that God might become irritated as the Holy Word was conveyed without the human voice. Remember that "Per Sona" became later the equivalent of "Person". In other words: the 'frozen' word became portable and detached from the person-to-person interaction. Later, the same happened to the strip book, where chunks of text were captured in drawings; also, here the irrevocable unity between the spoken and the printed word became broken and caused significant social resistance.
71. The entry of telephony, radio, and television polarized society, even for religious reasons; "who had the right to counterfeit the human senses?" In large parts of Western society, the new media were seen as sabotage and cheating the unity of persons' trust as it was no longer sure if exactly these words or human gesture were uttered.
72. Quite recently, the social media, fake news and the deep fake entered our daily mental 'territories' and are seen as potentially devastating for mutual trust and social cohesion.
73. AI as driver behind deep fake and state propaganda has a similar threat to the common opinion; similar to the magnitude of societal indignation in all three earlier media

40.2 References

40.2.1

1. Bhutani, A. and Wadhvani P. 2018. Artificial Intelligence (AI) in Education Market Size, By Model (Learner, Pedagogical, Domain), By Deployment (On-Premise, Cloud), By Technology (Machine Learning, Deep Learning, Natural Language Processing (NLP)), By Application (Learning Platform & Virtual Facilitators, Intelligent Tutoring System (ITS), Smart Content, Fraud & Risk Management), By End-Use (Higher Education, K-12 Education, Corporate Learning), Industry Analysis Report, Regional Outlook, Growth Potential Competitive Market Share & Forecast, 2018 – 2024. Available at: <https://www.gminsights.com/industry-analysis/artificial-intelligence-ai-in-education-market> (Accessed 29 December 2022).
2. Dong, X., Wu, J. and Zhou, L. 2017. Demystifying AlphaGo Zero as AlphaGo GAN. Available at: <http://arxiv.org/abs/1711.09091> (Accessed 15 February 2022).
3. Holmes, W., Bialik, M. and Fadel, C. 2019. Artificial Intelligence in Education: Promises and implications for teaching and learning. Boston, MA, Centre for Curriculum Redesign.
4. Hume, K.H., 2017. Artificial intelligence is the future—but it’s not immune to human bias. Macleans. Available at: <https://macleans.ca/opinion/artificial-intelligence-is-the-future-but-its-not-immune-to-human-bias/> (Accessed 28 March 2022).
5. Hao, K. 2019. In 2020, let’s stop AI ethics-washing and actually do something - MIT Technology Review. MIT Technology Review. Available at: <https://www.technologyreview.com/s/614992/ai-ethics-washing-time-to-act/> (Accessed 13 January 2022).
6. Smith, M. L. and Neupane, S. 2018. Artificial Intelligence and Human Development. Toward a Research Agenda., Ottawa, International Development Research Centre. Available at: <https://www.idrc.ca/en/stories/artificial-intelligence-andhuman-development> (Accessed 22 February 2021).
7. Marcus, G. and Davis, E. 2019. Rebooting AI: Building artificial intelligence we can trust. New York, Ballantine Books Inc.
8. Lucas, L. 2018. China’s artificial intelligence ambitions hit hurdles. Financial Times. Available at: <https://www.ft.com/content/8620933a-e0c5-11e8-a6e5-792428919cee> (Accessed 17 February 2022).
9. Lupton, D. and Williamson, B. 2017. The datified child: The dataveillance of children and implications for their rights’, *New Media & Society*, Vol. 19, No. 5, pp. 780–794.
10. Mitchell, M. 2019. Artificial Intelligence: A guide for thinking humans. London, Penguin.
11. Moravec, H. 1988. Mind Children: The future of robot and human intelligence. Boston, MA, Harvard University Press.



PRISCILLA



priscilla.fitped.eu