The Age of Artificial Intelligence

An Exploration

Edited by
Steven S. Gouveia
University of Minho, Portugal

Cognitive Science and Psychology
Vernon Press

www.vernonpress.com

In the Americas:
Vernon Press
1000 N West Street,
Suite 1200, Wilmington,
Delaware 19801
United States

In the rest of the world:
Vernon Press
C/Sancti Espiritu 17,
Malaga, 29006
Spain

Cognitive Science and Psychology

Library of Congress Control Number: 2020931461

ISBN: 978-1-62273-872-4

Cover design by Vernon Press using elements designed by FreePik.

TABLE OF CONTENTS

List of Acronyms

ACM = Agent cognitive model

ADM = Agent dialogue model

AGI = Artificial General Intelligence

AI = Artificial Intelligence

AIML = Artificial intelligence markup language

AISE = Artificial intelligence safety engineering

ALAMO = Atelier de Littérature Assistée par la Mathématique et les Ordinateurs

ALICE = Artificial linguistic internet computer entity

APMA = Amino-3-hydroxy-5-methyl-4-isoxazole propionic acid

AKM = Agent knowledge model

A.R.T.A = Atelier de Recherches et Techniques Avancées

BCI = Brain-computer interfaces

BDI = Belief–desire–intention model

BNC = British national corpus

BSC = Bodily self-consciousness

CSs = Constructional schemes

CD = Compact disc

CERN = European organization for nuclear research

CEV = Coherent extrapolated volition

CGs = Conceptual graphs

CL = Computational linguistics

CNS = Central nervous system

COGUI = Conceptual graphs user interface

CP = Cognitive phenomenology

CREB = cAMP response element-binding protein

CRISPR = Clustered regularly interspaced short palindromic repeats

CS = Computational structures

CSA = Conversational software agent

DARPA = Defense advanced research projects agency

DDD = Distributed, decentralized, and democratic

DLT = Distributed ledger technology

DM = Discourse management

DNA = deoxyribonucleic acid

DIY = Do it yourself

DRM = Digital rights management

DRT = Discourse representation theory

ELO = Electronic literature organization

EMI = Experiments in musical intelligence

EQ = Embodiment question

FAI = Friendly artificial intelligence

FEP = Free-energy principle

FOOM = Recursively self-improving artificial intelligence engendered singularity

FOS = Free and open source

GANs = Generative adversarial networks

GM = Grand masters (of chess)

GOFAI = Good old-fashioned artificial intelligence

GPS = Global positioning system

HIS = Hazardous intelligent software

HOT = higher-order thought theory of consciousness

HS = Hazardous software

IBM = International business machines

IEEE = Institute of electrical and electronics engineers

IQ = Intelligence quotient

IoT = Internet of things

KL = Kullback-Leiber

KR = Knowledge representation

LS = Logical structure

LSc = Layered structure of the clause

LSDB = Liveset database

MAI = Malevolent artificial intelligence

MEG = Magnetoelectroencephalography

MIRI = Machine intelligence research institute

MS = Mental states

MT = Machine translation

MW = Musical work

NL = Natural language

NLP = Natural language processing

NLU = Natural-language understanding

NGO = Non-governmental organization

NOC = Non-ordinary consciousness

NCC = Neural correlates of consciousness

NSA = National security agency

OR = Ockham's razor

Orch-OR = Orchestrated objective reduction

PCR = Polymerase chain reaction

PHEN = Phenomenological independence implies metaphysical independence

POS = Part-of-speech

QHOT = Quotational higher-order thought theory of consciousness

R&D = Research and development

RDF = Resource description framework

RNA = Ribonucleic acid

RRG = Role and reference grammar

SA = Simulation argument

SAC = Speech act constructions

SAT = Speech act theory

SH = Simulation hypothesis

SP = Sensory processing

STAGE = Science, technology and governance in europe

TOC = Theatre of consciousness

TOTE = Test-operate-test-exit

TREC = Text retrieval conferences

TT = Turing test

WSD = Word sense disambiguation

XML = Extensible markup language

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

“Computers are useless because they can only give you answers”
(Pablo Picasso)[1]

"Chess... is eminently and emphatically the philosopher's game”
(Paul Morphy)[2]

On May 11th, 1997, in the emblematic city of New York, on the 9th floor of the Equitable Center, a major event happened that can be considered as one of the most decisive moments for humanity:[3]

> 1.e4 c6 2.d4 d5 3.Cc3 dxe4 4.Cxe4 Cd7 5.Cg5 Cgf6 6.Ad3 e6 7.C1f3 h6 8.Cxe6 De7 9.O-O fxe6 10.Ag6+ Rd8 11.Af4 b5 12.a4 Ab7 13.Te1 Cd5 14.Ag3 Rc8 15.axb5 cxb5 16.Dd3 Ac6 17.Af5 exf5 18.Txe7 Axe7 19.c4 1-0.

If you don’t know the history of some of the most famous chess games, you will not immediately recognize this sequence. But if you do, you will notice that this is the sum-up of all the 19 moves made in the sixth game of the second match between Garry Kasparov (one of the most brilliant chess players in the world - alongside M. Carlsen, B. Fisher, V. Anand, A. Karpov and P. Morphy) and the computer created by International Business Machines (IBM) named DeepBlue. In the 8th move, after Kasparov played h6 – something he would regret later – DeepBlue decided to sacrifice a knight for a pawn, playing e6, something that Kasparov would never have expected.

What happened next was the pure dominance of the computer, forcing Kasparov to resign on the 19th move and to lose the match.[4] Kasparov would

---

[1] Different versions of this are cited in William Fifield’s original interview with Picasso, “Pablo Picasso: A Composite Interview,” published in the Paris Review 32, Summer–Fall 1964and in Fifield’s 1982 book *Search of Genius*, New York: William Morrow.

[2] As quoted in *Testimonials to Paul Morphy*, Presented at University Hall, New York, May 25, 1859 (cf. https://play.google.com/books/reader?id=aEZAAAAAYAAJ).

[3] *Time* magazine had claimed something similar regarding the first game of the first match between Kasparov and DeepBlue in 1996. Kasparov had lost the first game (although he would ended up winning the match): the first-game defeat was more than “world historical. It was species-defining”.

[4] The match was composed by six gamesand the result after the first five games was a tie, 2½–2½. The match’s result was, after the win of DeepBlue in the sixth and last game, 3½–2½.

never have played h6 against a human opponent, nevertheless, since he was playing against a computer, he chose that move: “I didn't want to play. I was sorry about my decision to play h6. Normally computers don't take on e6”.

Although DeepBlue was built with 256 co-processors capable of calculating approximately 200 million positions per second, for Kasparov, the reason for his loss was simple: no computer would use a tactical move such as sacrificing a knight so early in the game. At the time, there were some suspicions that the research team behind the development of DeepBlue was being helped (live in action) by Grand Masters (GM), namely, Bobby Fisher.

In the late eighteenth century, there was a famous chess-playing mechanical automaton called “the Turk[5]”. It was an engraved figure made of wood that could move its pieces and it could play a competent chess game. The “Turk” was branded as the very first artificial system. Touring through America and Europe, it played against professional players, including renowned historical celebrities such as Napoleon Bonaparte and Benjamin Franklin, who were themselves chess aficionados. Of course, this was a hoax, an elaborate, but nevertheless, a fake artificial system with a person cleverly hidden inside the wood structure playing all the moves (cf. Kasparov, 2017: 7).

IBM had one purpose only: to prove that they could build a machine that could defeat the best chess player in the world, the reigning champion. The team responsible for the development of DeepBlue was composed by Murray Campbell (IBM, Thomas J. Watson Research Center), Joel Benjamin (Chess GM and consultant), Feng-hiung Hsu (who started developing DeepBlue while he was at Carnegie Mellon University), Thomas Anantharaman as well as a few others, all managed by Chung-Jen Tan, who was known as the spokesman and the “resident philosopher” of the team.

Kasparov had won the first match against DeepBlue in the previous year, 15 months prior in Philadelphia - this was the second one. The preparation for this second match was difficult because he couldn’t study previous games played by DeepBlue since there were none: he had to play against a black box without any chance of studying and analyzing previous games made by the computer. Worse than that, there was a clause in the contract for the second match that Kasparov completely overlooked: the machine could be rebooted during or after each game. This would make the post-analysis of its games impossible, therefore, eliminating any chances of studying specific games or to recognize any patterns in DeepBlue’s approach to chess.

[5] The machine was nicknamed the Turk because it played its moves through a turbaned marionette attached to a cabinet (cf. Campbell, 1997: 83).

Newsweek's cover called this match "The Brain's Last Stand". The match was covered in all newspapers and broadcasted live on television: the world was seeing firsthand that artificial machines could potentially surpass a game in which human beings had excelled for centuries. After this first conquest, Artificial Intelligence research focused on building computers that could defeat humans in other more complex games, like Go (with AlphaGo), Scrabble (with Quackle) or Jeopardy (with Watson). The artificial system won in every one of these games.

The 1997 match was announced as Kasparov representing Humanity versus the Machine. If he lost the match, then everything could be achievable for AI: it would officially give rise to the "Age of Artificial Intelligence". For contemporary science, chess was seen as the ultimate test of intelligence. In cinema, there are also many chess references along these lines: for example, in Stanley Kubrick's *2001: a Space Odyssey*, there is a scene where the computer HAL9000 plays a chess game against Frank Poole - the game is an actual recreation of a tournament game from 1913 between A. Roesch and W. Schlage (cf. Campbell, 1997: 79).

It's very interesting to notice that the birth of Computer Science and AI is associated with the first reflections and thoughts about creating a machine that could play chess. One of the first discussions happens in Charles Babbage's *The Life of a Philosopher* (1845). In 1945, one century later, Konrad Zuse describes a program that could generate legal moves in chess. Claude Shannon, founder of information theory, wrote a paper titled "A Chess-Playing Machine" (1950) where he describes two main approaches in building a competent chess-program: Type A was about creating a program based on brute-force, that is, a program that could analyze and calculate in seconds all the millions of potential positions for a specific move; and Type B, a program that would have a strategic and goal-focused approach, more like humans' beings play chess. Shannon believed that the Type B approaches would be more successful in beating a human, but it was eventually the Type A approach which won the race, culminating in the development of DeepBlue.

Alan Turing also thought about whether an artificial system that could play chess, called "Turochamp", famously known as the "Turing's paper machine". The program, created in 1952,[6] was written with his colleague David Champernowne and was composed by a specific set of instructions. The only problem is that there were no digital computers in that time, so Turing wrote

---

[6] See the original publication here: see https://en.chessbase.com/post/reconstructing-turing-s-paper-machine).

the program by hand - it was later recreated in an actual computer, where it turned out to not be a very competent chess player.

In 1950, one of the founders of Artificial Intelligence, Herbet Simon, stated it would only take 10 years for a machine to become the world champion of chess – he was just mistaken by 30 years (cf. Cambpell, 1997: 83-85).

In 1956, John McCarthy, owning to Alexander Kronrod, described chess as the "drosophila of Artificial Intelligence" (cf. McCarthy, 1990). Like the common fruit fly, which juxtaposed to more complex organic systems, its research is quite "simple" to do. Nevertheless, it can produce significant knowledge about further complex systems. Just as at its base, chess is a "simple" game, and therefore, it can teach us an ample amount about human cognition and intelligence (cf. Ensmenger, 2012).

As we can see, there was a great, almost obsessive focus on chess in the first days of AI research because it was believed that chess was the ultimate pinnacle of human intelligence. Nowadays, this seems, a tiny bit exaggerated. At present, Artificial Intelligence is focused on what is called "weak AI": it excels at very specific tasks – like playing games, facial recognition or driverless cars – but it is not even close to achieving human-level intelligence. The reason is quite simple: Artificial Intelligence's research methods are more about imitating human performances – the Turing Test is a very good example of this idea – than to look for its own achievements and goals.

Consequently, all the tasks we can describe and codify can be outperformed by machines. But the real achievement of a fully conscious machine seems far still. Because we do not know anything about consciousness, Artificial Intelligence conceived as "strong AI", that is, a conscious A.I., may never be fully achieved. For that, we need the right theoretical framework – we need better and more philosophical research.

The book is divided into five main sections. Section I is dedicated to reflections on the Intelligence of AI and will open with a chapter by Soenke Ziesche and Roman V. Yampolskiy, which discuss the mathematics of intelligence for grouped minds, nested minds as well as deducted minds. The following chapter, by Stevan Harnad, debates if the existence of feelings is a real caveat for a system that would pass the Turing's Test. Next, Daniel Dennett argues against the mysterianism position that we cannot study our conscious mind and explain why AI, although theoretically possible to be achieved, may never be practically accomplished because of its costs and the lack of epistemic advantages of such an achievement. Finally, closing Section I, David Pearce discusses three different ways to connect human and artificial intelligence: by fusion, replacement or co-evolution, arguing that only the third process may be plausible.

Section II follows, dedicated to discussion on the relationship between consciousness, simulation and artificial intelligence. Gabriel Axel Montes and Ben Goertzel present the concept of a 'mindplex' as a way of enhancing the connection between human and artificial minds; for which, they use the concept of non-ordinary consciousness (NOC) and show how that perspective can be relevant for understanding the mind and cognition in general. Cody Turner follows offering two arguments in favor of the thesis that a phenomenology of cognition is neither reducible to, nor dependent upon, sensory phenomenology. If this thesis is plausible, then it follows that AI consciousness may not require embodiment to be emulated, as commonly assumed. The next chapter by Nicole Hall argues that aesthetic experience is a fundamental feature of human consciousness and separates human from artificial intelligence. She argues further that it is a mistake to confuse the mere possibility of achieving "conscious singularity", as she defines it, with human consciousness and its capacity for aesthetically experiencing natural environments. To conclude Section II, Steven S. Gouveia introduces an intriguing idea that we may be living in a computer simulation, briefly debating the main reasons in favor and against this hypothesis.

Section III, dedicated to aesthetical creativity and language in artificial intelligence, opens with a chapter by Caterina Moruzzi where in the light of recent developments in AI music software generators discusses the question, "Can a computer create a musical work?" On the same topic, René Mogensen proposes a formal representation of content in computational creativity of music, noting that in order to achieve complete computational music creativity, aesthetic experience appears to be necessary. Mariana Chinellato Ferreira follows, applying the same discussion about aesthetical creativity in computer-generated literature, analyzing specific software such as PoeTryMe. Closing Section III, Kulvinder Panesar presents a functional linguistic perspective on natural language processing in artificial intelligence.

The subsequent Section IV is on the Ethics of the Bionic Brain Peter A. DePergola II opens by offering the argument that neurocognitive enhancement can be justified as morally plausible if it (a) promotes general moral character, (b) complements human nature and (c) effects a deeper sense of individual and social identity. Next, Tomislav Miletić and Frederic Gilbert discuss the potential harms of brain-computer interfaces (BCI) on self-determination, warning that any patient who accepts the use of such future AI medical technology should be sufficiently prepared for the symbiotic relation before the implementation. Following on the same topic of the ethics of BCI, Aníbal M. Astobiza, Txetxu Ausin, Ricardo M. Ferrer and Stephen Rainey focus on some issues raised by BCI research, identifying some dangers, challenges and opportunities for the elaboration of a common ethical and legal framework concerning issues of safety, ethics and data protection. To conclude Section IV, Natasha Vita-More argues that Artificial General Intelligence (AGI) can be used as a tool to improve our knowledge about ourselves and the world.

Finally, to close the book, Section V follows on the Ethics of Artificial Intelligence, starting with a chapter by Federico Pistono and Roman V. Yampolskiy that provides some general guidelines for the creation of a Malevolent Artificial Intelligence (MAI) with the goal of challenging the AI Safety Community to continue its effort by discovering and reporting specific problems. Following, Hasse Hämäläinen attempts to find the most plausible answer to the question of whether a machine could be attributed moral and legal rights and obligations, arguing that if a machine can perform a specific task or set of capacities as human beings do, then the rights and obligations of humans should also be applied to machines. The next chapter by Vernon Vinge discusses the ethical implications of the Singularity, offering a set of ethical guidelines to avoid the extinction of human race. Finally, to conclude both Section V and the book, Eray Özkural discusses the ethical and epistemological implications of the Free Energy Principle: the idea that a self-organization occurs through minimization of free energy.

The Age of Artificial Intelligence is imminent, if it's not already here. We should make sure that we invest in the right people and the right ideas in order to create the best solutions possible. My hope is that this book will help to influence the right minds. If Reason killed god in the 20th and 21st century, Reason – philosophy, science and technology – may resurrect it in form of an Artificial General Intelligence: an AI that may know everything about anything. We should make sure that we create the right kind of god and that we keep it in the right hands.

I would like to finish this introduction by deeply thanking all the people who made this project feasible.[7]

Steven S. Gouveia
Ottawa, 10/09/2019

[7] I would like to acknowledge the precious help offered by Jessica Clarke and, to Susan Schneider for her valuable feedback.

References

Campbell, Murray S. (1997) "'An Enjoyable Game': How HAL Plays Chess" In *HAL's Legacy: 2001's Computer as Dream and Realtiy* (David G. Stock, ed.), Cambridge, MA: MIT Press.

Ensmenger, Nathan (2012) "Is chess the drosophila of artificial intelligence? A social history of an algorithm" *Soc Stud Sci.,* 42 (1): 5-30.

Feng-hsiung, Hsu, Anantharaman, Thomas, Campbell, Murray and Nowatzyk, Andreas (1990) "A Grandmaster-level Chess Machine" *Scientific American*, 263 (4): 44-51.

Fifield, William (1982) *Search of Genius*, New York: William Morrow

Frey, Peter W. (ed.) (1983) *Chess Skill in Man and Machine*, New York: Springer-Verlag.

Hsu, Feng-hsiung (2002) *Behind Deep Blue: Building the Computer that Defeated the World Chess Champion*, Princeton, NJ: Princeton University Press.

Kasparov, Gary (2017) *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*, New York: PublicAffairs™.

Levy, David and Newborn, Monty (1991) *How Computers Play Chess*, New York, NY: Computer Science Press.

Marsland, Anthony T. and Schaeffer, Jonathan (eds.) (1990) *Computers, Chess and Cognition*, New York, NY: Springer-Verlag.

McCarthy, John (1990) "Chess as the Drosophila of A.I.". In *Computers, Chess and Cognition* (Marsland T.A., Schaeffer J., eds.), New York, NY: Springer, pp. 227- 237.

Shannon, Claude (1950) "Programming a computer for playing chess" *Philosophical Magazine*, 41: 256-75.

PAGES MISSING

FROM THIS FREE SAMPLE

Contributors

Aníbal Monasterio Astobiza is a Basque Government Posdoctoral Researcher. His research lies at the intersection of the cognitive, biological and social sciences exploring their philosophical underpinnings. He is a member of international (and national) research projects such as for example, EXTEND and INBOTS (Horizon 2020). He explores topics such as roboethics, machine ethics, ethics of AI, data ethics, ethics of neurotechnology, human enhancement, space ethics, applied ethics, bioethics, philosophy of neuroscience, philosophy of psychiatry, altruism, cooperation or philosophy of network science and complexity. Aníbal graduated in Philosophy (Universidad de Deusto) and obtained his PhD in Cognitive Science and Humanities at the Universidad del País Vasco/Euskal Herriko Unibertsitatea with a dissertation on social cognition. Monasterio Astobiza A. et al. (2019), Bringing inclusivity to robotics with INBOTS. Nature Machine Intelligence 1, 164.

Ben Goertzel is the CEO of the decentralized AI network SingularityNET, a blockchain-based AI platform company, and the Chief Science Advisor of Hanson Robotics, where for several years he led the team developing the AI software for the Sophia robot. Dr. Goertzel also serves as Chairman of the Artificial General Intelligence Society, the OpenCog Foundation, the Decentralized AI Alliance and the futurist nonprofit Humanity+. Dr. Goertzel is one of the world's foremost experts in Artificial General Intelligence, a subfield of AI oriented toward creating thinking machines with general cognitive capability at the human level and beyond. He also has decades of expertise applying AI to practical problems in areas ranging from natural language processing and data mining to robotics, video gaming, national security and bioinformatics. He has published 20 scientific books and 140+ scientific research papers, and is the main architect and designer of the OpenCog system and associated design for human-level general intelligence.

Caterina Moruzzi is a postdoctoral researcher at the University of Nottingham where she obtained her PhD in 2018 with a thesis on the ontology of musical works. She has a MA in Philosophy at the Università di Bologna and she graduated in Piano performance from the Conservatorio G.B. Martini in Bologna. She is actively contributing to research in aesthetics and Philosophy of Art by presenting at conferences worldwide and by publishing on high-impact journals. In her current research she investigates the impact of revolutionary development of AI technology on ontological, aesthetic, and evaluative considerations on music. In particular, she aims to answer two research questions: 'How does the development of computer-generated music affect the ontology of musical works?' and 'Do we deem computer-generated music as creative as human-generated music? If not, why?'. The scope of this research is to fill the existing gap in the literature in regard to the impact that cutting-edge developments in AI have on aesthetic and ontological debates on music and art.

Cody Turner is a third year PhD student in philosophy at the University of Connecticut. His areas of research include the philosophy of consciousness, the philosophy of artificial intelligence, and applied ethics. He is a member of the AI, Mind, and Society Group at the University of Connecticut, and has a teaching appointment with the Computer Science & Engineering department. He currently has one other publication entitled 'Could you Merge with AI? Reflections on the Singularity and Radical Brain Enhancement' (with Susan Schneider, forthcoming in Oxford Handbook of Ethics of AI, Oxford University Press, 2020).

Daniel C. Dennett was educated at Harvard (BA 1963) and Oxford (DPhil, 1965) and taught at UC Irvine (1965-71) before moving to Tufts University, where he has taught ever since. He is currently University Professor and Austin B. Fletcher Professor of Philosophy. He is the author of many books, including "Consciousness Explained" (1991), "Darwin's Dangerous Idea" (1995) and "From Bacteria to Bach and Back" (2017) and hundreds of articles.

David Pearce is author of The Hedonistic Imperative (1995), which advocates using biotechnology to abolish suffering throughout the living world in favour of gradients of superhuman bliss. In 1998, Pearce co-founded with Nick Bostrom The World Transhumanist Association, now rebranded as Humanity Plus. Transhumanists urge the use of technology to create a "triple S" civilisation of superhappiness, superlongevity and superintelligence.

Eray Özkural has obtained his PhD in computer engineering from Bilkent University, Ankara. He has a deep and long-running interest in human-level AI. His name appears in the acknowledgements of Marvin Minsky's The Emotion Machine. He has collaborated briefly with the founder of algorithmic information theory Ray Solomonoff, and in response to a challenge he posed, invented Heuristic Algorithmic Memory, which is a long-term memory design for general-purpose machine learning. Some other researchers have been inspired by HAM and call the approach "Bayesian Program Learning". He has designed a next-generation general-purpose machine learning architecture. He is the recipient of 2015 Kurzweil Best AGI Idea Award for his theoretical contribution to universal induction. He has previously invented an FPGA virtualization scheme for Global Supercomputing, Inc. which was internationally patented. He has also proposed a cryptocurrency called

Cypher, and an energy based currency which can drive green energy proliferation. You may find his blog at http://log.examachine.net and some of his free software projects at https://github.com/examachine.

Federico Pistono is an Entrepreneur, Angel Investor, Researcher, and Educator. He's author of the international success "Robots Will Steal Your Job, But That's OK", has co-founded four startups, he's Futurist and Technology Expert in the TV show "Codice: la vita è digitale" (Code: life is digital), which airs on Rai1, and is Former Head of Blockchain of HyperloopTransportation Technologies. He's founder and CEO of Exponential Thinking, which invests in early stage startups, and creator ofMavericks, a school for Entrepreneurs, Investors, and Innovators. He holds a BSc in Computer Science from the University of Verona, and in 2012 he graduated from Singularity University, NASA Ames Research Park.

Frederic Gilbert currently research consists in exploring the ethics of novel implantable brain-computer interfaces operated by Artificial Intelligence. At the time of writing this bio, he is an Australian Research Council (ARC) Discovery Early Career Research Award recipient and Senior Lecturer in Ethics at CALE, School of Humanities, at the University of Tasmania. He is affiliated with the Ethics, Policy & Public Engagement program of the ARC Australian Centre of Excellence for Electromaterials Science (ACES), and concomitantly an Ethics Consultant for the Centre for Neurotechnology, for which he conducts research at the University of Washington, in Seattle, USA. He has published over 65 articles in bioethics and neuroethics. Some of his main publications related to this chapter are: Gilbert F., O'Brien T., Cook, M. (2018) “The Effects of Closed-Loop Brain Implants on Autonomy and Deliberation: What are the Risks of Being Kept in the Loop?” In Cambridge Quarterly of Healthcare Ethics, pp. 1-12; Gilbert F., Cook, M., O'Brien, T., Illes J. (2019) “Embodiment and estrangement: Results from a first-in-human intelligent BCI trial” in Science and Engineering Ethics, 25 (1): 83-96; Gilbert, F. (2015) “A threat to Autonomy? The Intrusion of Predictive Brain Devices”. American Journal of Bioethics Neuroscience, 6 (4): 4-11.

Gabriel Axel Montes, Ph.D. Candidate, is a neuroscientist, educator, and musician. His 12+ years of research in cognitive neurobiology and mind-body practice has elucidated both endogenous and technological mechanisms for intervening in the brain’s process of modelling reality and perception. His recent research focuses on informing the development of beneficial artificial general intelligence (AGI) with insights from the non-ordinary consciousness (NOC) of mind-body practices. Gabriel is known for fusing insights from rigorous NOC self-practice into the design of AGI, and virtual/augmented reality and humanitarian projects involving resilience education. Gabriel is an international educator on the intersection of neuroscience and mind-body practice. As

Founder of Neural Axis®, he leads and consults on neuroscience-based technology design, strategy, and behavioural change. He leads organisations, communities, and individuals in applying neuroscience to first-person experience. For more about Gabriel, please visit gabrielaxel.com.

Hasse Hämläinen received his PhD in Philosophy from the University of Edinburgh (2015). At the time of writing, he is working as a QA analyst for a leading IT company and as a project researcher at the Jagiellonian University in Kraków. His research is focused on the history of moral theory, especially on ancient and the 18th century moral thinking, and, recently, also on charting the ethical challenges and limits of AI technologies. His publications include a co-edited monograph, The Sources of Secularism: Enlightenment and Beyond (Palgrave Macmillan, 2017) and his work has been published, for example, in International Philosophical Quarterly, Diametros, Journal of Ancient Philosophy and in The Philosophical Forum.

Kulvinder Panesar is a strategically focused senior computing professional and an academic for over twenty years. Her current role is a Senior Lecturer of Computer Science at the School of Art, Design and Computer Science at York St John University, York in United Kingdom. Her research interest is NLP (Natural Language Processing) in AI (Artificial Intelligence), and meaning and knowledge representation in conversational software agents (CSAs), and more recently, Chatbots and conversational AI. This research area is multi-disciplinary spanning agent thinking, linguistics, and computational linguistics, natural language processing (NLP), knowledge representation (KR), Semantic Web (SW), artificial intelligence (AI) and data science. Current teaching commitments include undergraduate and postgraduate subjects in: computer science, software engineering design and development, website design and development, advanced database and networks; programming, professional practice; project management, project supervision, mathematics for computer science, philosophies of technology, and artificial intelligence. Previous teaching undertaken include: management information systems, marketing and e-commerce strategy (customer relationship management) and development, enterprise computing solutions, systems analysis and design, formal methods and research methods. Additional roles undertaken over the years are course director, module leader, final year project coordinator and supervisor, placement co-ordinator, 'study abroad' academic advisor, researcher, internal verifier, mentor, STEM ambassador, external assistant examiner (EdExcel), ICT teacher, IT support/programmer/trainer contractor, and systems analyst. She is a professional member (MBCS) of the British Computer Society (BCS), and a Fellow of the Higher Education Academy (AdvanceHE) and currently organise and co-ordinate the Women in Computer Science (WICS) group at York St John University.

Mariana Chinellato Ferreira is a Brazilian researcher, graduate at Social Communication with specialization in Advertising and Publicity from the Catholic University of Santos (2002) and graduate at Modern Languages, major in Portuguese and English languages and Literature from the Catholic University of Santos (2009). Has an exchange program period at the University of Coimbra (Portugal - 2007/2008) in Modern Languages. Has professional experience in Linguistics, focusing on Applied Linguistics, acting on the following subjects: Modern Foreign Language (English), English for Specific Purposes, Portuguese as a Second Language, Translation (Portuguese – English) and Education. Professional academic experience in English and American Literature, Portuguese and Brazilian Literature, Theory of Literature, Theater and Poetry. She has a Master's degree in Architecture at University of São Paulo - São Carlos with the project: “City and Literary form: Urban representation in the Brazilian contemporary literature”. Currently, she is a PhD Candidate of the PhD Program in Materialities of Literature at the University of Coimbra with a research about questions of language and creativity in Narrative Generation Systems in a partnership with joint supervision of the Centre for Informatics and Systems of the University of Coimbra where she develops a new model of narrative in order to apply into a prototype system of automatic narrative generation in Portuguese language. Her last publish article discusses new models of narrative in computer-generated literature: “Seeking the Ideal Narrative Model for Computer-Generated Narratives”, presented at the CC-NLG 2018, in Tilburg, the Netherlands.

Natasha Vita-More is an author, humanitarian, and innovator whose work focuses on longevity and regenerative generations. As a motivational speaker, she focuses on causes and solutions, while fostering meaningful acknowledgement of others' works who have aspired to identify human potential. She is called “An early adapter of revolutionary changes” (Wired, 2000) and “Advocates the ethical use of technology to expand human capacities” (Politico, 2017). Natasha was the Lead Scientific Researcher on the Memory Project, which scientific breakthrough concerns long-term memory of C. elegans and cryonics (2015). As a proponent for mitigating aging, Natasha introduced the seminal field of Human Enhancement for longevity in academics. Her expertise in the field of ethics has produced high-level scholarship for undergrad and graduate students in the fields of computer science, cybersecurity, robotics, gaming, and business fields. As an entrepreneur, her experience within the domain of foresight studies has established principles and practices for assessing humanity's potential futures. Natasha originated the original “Whole-Body prosthetic”, a seminal innovation comprised of nanomedicine, AI, and robotics that spearheaded alternatives to biological aging and received international recognition. Currently, she is a Professor of Humanities in Ethics, Business, and Entrepreneurship, and serves as the Executive Director of Humanity+. She is also a Fellow at the Institute for Ethics and Emerging Technologies, and author of numerous academic articles and books. Some publications: "Persistence of Long-Term Memory in Vitrified and Revived Caenorhabditis elegans" or "The Transhumanist Reader". More can be find here: https://natashavita-more.com/publications/.

Nicole Hall most recently held a Fernand Braudel research fellowship at the Institut Jean Nicod (Ecole Normale Supérieure) and taught aesthetics at the Université Paris 8. Her PhD, on the nature of aesthetic experience, is from the University of Edinburgh. Her research is at the intersection of aesthetics, environmental philosophy, the philosophy of mind and consciousness, and cognitive and neuroscience. She has a forthcoming publication on Adam Smith's aesthetic psychology and is working on the role of perception in aesthetic experience.

Peter DePergola II is Director of Clinical Ethics at Baystate Health, Assistant Professor of Medicine at University of Massachusetts Medical School, and Assistant Professor of Bioethics and Medical Humanities at Elms College, where he serves concurrently as Director of the Center for Ethics, Religion, and Culture. Dr. DePergola earned his B.A. degree in philosophy and religious studies at Elms College, his M.T.S. degree in ethics at Boston College, and his Ph.D. degree in healthcare ethics at Duquesne University. He completed his residency in neuroethics at University of Pittsburgh School of Medicine, his fellowship in neuropsychiatric ethics at Tufts University School of Medicine, and his advanced training in neurothanatological ethics at Harvard Medical School. Dr. DePergola's recent book, Forget Me Not: The Neuroethical Case Against Memory Manipulation(Vernon Press, 2018), has been critically acclaimed as a landmark achievement in the field of neuroethics, and was the #1 New Release in Medical Ethics on Amazon.com at the time of its publication. The recipient of numerous awards for academic scholarship and healthcare leadership, his current research explores the theoretical and empirical metaphysics of hope, particularly as it relates to the neuroethics of moral virtue and end-of-life decision making.

René Mogensen holds a PhD from Birmingham City University/Royal Birmingham Conservatoire in the UK, an MM from the Royal academy of Music in Aarhus, Denmark, an MA from New York University and a BA from the University of Rochester, NY, USA. His research in music, computer music, and artificial intelligence in music has recently been published in books by Springer, Éditions Delatour, and as articles in the Journal of Creative Music Systems, the Early Music journal, as well as in international conference proceedings (CSMC, AISB, EuroMAC). He is the recipient of grants from the Danish Arts Foundation, Meet the Composer USA, John Anson Kittredge Fund, Yvar Mikhashoff Trust for New Music, and other funders. His numerous original music works for ensembles and soloists (with and without computers) have been performed around Europe, the USA and Asia. He has participated as a performer/composer in many music concerts during the past three decades around the USA and Europe. He currently teaches undergraduates and postgraduates in areas of creative music technology as well as in music improvisation and composition at the Royal Birmingham Conservatoire in the UK.

Ricardo Morte Ferrer is a lawyer and Data Protection Consultant (https://sdm2.de/), he is also a certified ITIL-F and Lead Auditor ISO 27001. He is a member of the Working Group Digitalization and Health at the German Academy for Ethics in the Medicine. PhD candidate at the University of Granada working on"Ethical and Philosophical Aspects of Privacy". Nowadays is researcher in INBOTS CSA network: Inclusive Robotics for a better Society [www.inbots.eu/], EXTEND: Bi-directional Hyper-connected Neural System [www.extend-project.eu/] and member of the core team at GNU Health.

Roman V. Yampolskiy is a Tenured Associate Professor in the department of Computer Engineering and Computer Science at the Speed School of Engineering, University of Louisville. He is the founding and current director of the Cyber Security Lab and an author of many books including Artificial Superintelligence: a Futuristic Approach. During his tenure at UofL, Dr. Yampolskiy has been recognized as: Distinguished Teaching Professor, Professor of the Year, Faculty Favorite, Top 4 Faculty, Leader in Engineering Education, Top 10 of Online College Professor of the Year, and KAS Outstanding Early Career in Education award winner among many other honors and distinctions. Yampolskiy is a Senior member of IEEE and AGI, and former Research Advisor for MIRI and Associate of GCRI. Roman Yampolskiy holds a PhD degree from the Department of Computer Science and Engineering at the University at Buffalo. He was a recipient of a four year NSF (National Science Foundation) IGERT fellowship. Before beginning his doctoral studies Dr. Yampolskiy received a BS/MS (High Honors) combined degree in Computer Science from Rochester Institute of Technology, NY, USA. After completing his PhD dissertation Dr. Yampolskiy held a position of an Affiliate Academic at the Center for Advanced Spatial Analysis, University of London, College of London. Dr. Yampolskiy is an alumnus of Singularity University (GSP2012) and a Visiting Fellow of the Singularity Institute (Machine Intelligence Research Institute). Dr. Yampolskiy's main areas of interest are AI Safety, Artificial Intelligence, Behavioral Biometrics, Cybersecurity, and Genetic Algorithms.

Soenke Ziesche holds a PhD in Natural Sciences from the University of Hamburg, which he received within a Doctoral Program in Artificial Intelligence. He has worked since 2000 for the United Nations in the humanitarian, recovery and sustainable development fields as well as in data and information management. In addition to UN Headquarters in New York he has been posted to Palestine, Sri Lanka, Pakistan, Sudan, Libya, where he was temporarily the highest UN representative during the revolution, South Sudan, Bangladesh and Maldives. He is currently a consultant, mostly for the UN, for Disaster Information Management as well as AI. In this capacity he has served, on behalf of the UN, the Governments of Bangladesh, Maldives and Sri Lanka respectively. In 2017 he produced a 137-pages UN report on opportunities for big data, AI and emerging technologies in the implementation and review of the UN Sustainable Development Goals. His latest academic publications on AI safety are "AI & Global Governance: A Seat at the Negotiating Table for AI? Opportunities and Risks", "Potential Synergies Between The United Nations Sustainable Development Goals And The Value Loading Problem In Artificial Intelligence" and, together with Roman V. Yampolskiy, "Towards AI Welfare Science and Policies".

Stephen Rainey is a research fellow in the Oxford Uehiro Centre for Practical Ethics. He is a working in the Horizon2020-funded project BrainCom which is developing therapeutic brain-computer interfaces that will enable communication for users with debilitating speech conditions. Dr Rainey studied philosophy in Queen's University Belfast and obtained his PhD in 2008 with a thesis on rationality. He has taught a range of philosophical topics, and worked on a number of European Commission-funded research projects. These have included work on ethics, emerging technologies, and governance. He contributes to European Commission ethical and scientific evaluation panels for the funding of research projects. These have included proposals for ERC and Marie-Curie grants. Stephen has research interests in the philosophy of language, rationality, governance, and artificial intelligence.

Stevan Harnad's research interest is in cognitive science, open access and animal rights. He is currently professor of psychology at Université du Québec à Montréal (UQAM) and professor emeritus of cognitive science at the University of Southampton. Elected external member of the Hungarian Academy of Sciences in 2001 (resigned in protest, 8 October 2016), he was founder and editor-in-chief of the journal Behavioral and Brain Sciences 1978-2002 and Canada Research Chair in cognitive science 2001-2015. His research is on categorization, communication, cognition, and consciousness. He has written on categorical perception, symbol grounding, origin of language, lateralization, the Turing test, distributed cognition, scientometrics, and animal sentience. He is founder and editor-in-chief of the journal Animal Sentience.

Steven S. Gouveia is finishing his PhD at the University of Minho (Portugal), under the supervision of the philosopher Manuel Curado (University of Minho) and the neuroscientist Georg Northoff (University of Ottawa), funded by the Science and Technology Foundation. His primary focus of research is on the relationship between Neuroscience and Philosophy. He is a visiting researcher at the Minds, Brain Imaging and Neuroethics Group at the Royal Institute of Mental Health, Uni. Ottawa. He is a researcher of the Lisbon Mind & Reasoning Group (Nova University of Lisbon) and the Mind, Language and Action Group (Uni. Porto). He published 8 academic books e.g. (2018) "Perception, Cognition and Aesthetics" (eds.) (Routledge) or "Automata's Inner Movie: Science and Philosophy of Mind" (eds.) (Vernon Press) and, forthcoming, "The Science and Philosophy of Predictive Processing (eds.) (Bloomsbury). He is the host of an international documentary on AI with the participation of international researchers such as Peter Singer and Paul Thagard. He is finishing writing is new book "Homo Ignarus: Rational Ethics for an Irrational World". He was one of the speakers at the famous "Science of Consciousness Conference" (2019) and published several articles on AI, mind and brain, ethics and aesthetics. More: stevensgouveia.weebly.com.

Tomislav Miletić, born on May 7, 1984, in the city of Rijeka, Croatia. My personal and academic interests predominantly lie in exploring the ethical and social impacts of Artificial Intelligence inside the paradigm of Human Enhancement. I am especially focused on the prospect of symbiotic human-Ai systems, their hybrid moral-epistemic agency, status and system architecture and the relevant ethical questions concerning their design and use. Currently engaged in finishing my doctoral thesis on the subject of "Human-AI symbiosis: the possibility of moral augmentation". Relevant publications: Miletić T. "Extraterrestrial artificial intelligences and humanity's cosmic future: Answering the Fermi paradox through the construction of a Bracewell-Von Neumann AGI" in Journal of Evolution and Technology, Vol. 25:1, 2015, 56-73 or Miletić T. "Human Becoming: Cognitive and Moral Enhancement Inside the Imago Dei Narrative" in Theology And Science, Vol. 13:4, 2015, 425-445.

Txetxu Ausín is a Tenured Scientist at the Institute for Philosophy, CSIC (Spanish National Research Council) [www.ifs.csic.es]. He is the Director of the Applied Ethics Group (GEA). PhD in Philosophy (University of the Basque Country) and First Prize of the Year 2000. Invited Professor in several universities and researcher at the Institute for Democratic Governance Globernance. His research areas involve public ethics, bioethics, deontic logic, human rights, and philosophy of robotics and ICT. [Member of OECD Expert Group on Research ethics and new forms of data for social and economic research]. Editor and author of publications about this issues, is the Editor in Chief of the international journal DILEMATA on applied ethics [www.dilemata.net]. Nowadays is researcher in INBOTS CSA network: Inclusive Robotics for a better Society [www.inbots.eu/], EXTEND: Bi-directional Hyper-connected Neural System [www.extend-project.eu/] and BIODAT: Datos en salud[umucebes.es/]. Research Vice-President of ASAP-Spain (Academics Stand Against Poverty) and President of the Spanish Network of Philosophy [http://redfilosofia.es/]. Independent Member of the Public Ethics Commission of the Basque Government and Patron of the Foundation Cluster Ethics of the Basque Country and of Legal and Social Studies (EJYS Foundation), he also collaborates in the Committee of Experts on Aging of the Fundación General CSIC. Some publications: "Exclusion from healthcare in Spain: The responsibility for omission of due care". In Gaisbauer, Schweiger, Sedmak (eds.), Ethical Issues in Poverty Alleviation. Dordrecht: Springer, 2016. ; "Research Ethics and New Forms of Data for Social and Economic Research". OECD Science, Technology and Industry Policy Papers, No. 34, OECD Publishing, Paris.

Vernor Vinge has won five Hugo Awards, including one for each of his last three novels, A Fire Upon the Deep (1992), A Deepness in the Sky (1999), and Rainbow's End (2006). Known for his rigorous hard-science approach to his science fiction, he became an iconic figure among cybernetic scientists with the publication in 1981 of his novella "True Names," which is considered a seminal, visionary work of Internet fiction. His many books also include Marooned in Realtime and The Peace War. Born in Waukesha, Wisconsin and raised in Central Michigan, Vinge is the son of geographers. Fascinated by science and particularly computers from an early age, he has a Ph.D. in computer science, and taught mathematics and computer science at San Diego State University for thirty years. He has gained a great deal of attention both here and abroad for his theory of the coming machine intelligence Singularity. Sought widely as a speaker to both business and scientific groups, he lives in San Diego, California.

INDEX

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

R

S

T

U

W

Y

Z