Quantum Intelligence: Responsible Human-AI Entities

Melanie Swan [1], Renato P. dos Santos [2]

[1] *University College London, Malet Place, Gower Street, London, WC1E 6BT, UK*
[2] *Lutheran University of Brazil, Av. Farroupilha, 8001 - São José, Canoas, RS, 92425-020, Brazil*

Abstract

The increasing ability to harness quantum, classical, and relativistic scales, together with fast-paced change in generative AI and quantum computing, suggests the possibility of achieving not only technical but also social objectives, through responsible human-AI entities (intelligent agents interacting with competence and empathy). The social cannot be separated from the technical as intelligence may evolve into a platform-agnostic learning and problem-solving capability. A suite of concepts is introduced as *quantum intelligence*, *relativistic intelligence*, and *scale-free intelligence* to denote learning modes which incorporate scale-specific matter and space-time properties of quantum, classical, and relativistic domains. Modern technology has both risk and reward, and offers important theoretical and practical means of achieving socially beneficial outcomes. Theoretically, intelligence as the generic capability of agents (human or machine) to treat problems in multiple scale domains allows thinking to be reconceived as an operation of multiplicity, simultaneity, and portability, opening the scope of world to a wider sphere of concern beyond the traditional individual self and local community others. Practically, producing socially responsible AI could proceed in the phased approach of a "Moore's Law of AI Alignment" with short-term regulation and registries, medium-term internally learned reward functions, and long-term responsible human-AI entities. Humans are not naturally socially responsible but might be more so when operating with new levels of technological sophistication aimed at planetary-scale problems.

Keywords

AI alignment, socially responsible human-AI entities, AI math agents, quantum intelligence

1. Introduction

There is an imprecisely defined idea that AI should be socially responsible even though humans are not. Since technology advances faster than social maturity, the implication is to consider the trajectory of technological development as the first step in evaluating social impact. This work investigates how technological advance may facilitate the outcome of socially responsible human-AI entities.

The Oxford English Dictionary defines artificial intelligence as "the capacity of computers or other machines to exhibit or simulate intelligent behavior." On the one hand, AI is banal in continuing its multi-decadal rollout as a digital automation technology. On the other hand, the advent of generative AI (content produced by AI that is indistinguishable and possibly vastly more extensive than that produced by humans) may signal an inflection point that a qualitatively new era has been entered in AI that is puzzling, threatening, and exciting for humans.

Darwin wondered "Why should the brain be enclosed in a box?" [1], but AI advance hints that intelligence may no longer be constrained to the biological substrate. Generative AI is one of the world's most rapidly adopted technologies (the largest provider, OpenAI, reported 100 million ChatGPT users as of January 2023 [2]).

AAAI 2023 Spring Symposia, Socially Responsible AI for Well-being, March 27–29, 2023, USA
EMAIL: melanie@blockchainstudies.org (A. 1); renatopsantos@ulbra.edu.br (A. 2)
ORCID: 0000-0001-7739-4923 (A. 1); 0000-0003-1427-9973 (A. 2)

Although digital divide issues persist, the widespread worldwide accessibility and adoption of generative AI argues that socially beneficial objectives may already be in the process of being achieved as creative human minds explore and deploy the technology. Stanford's Global AI Vibrancy index cites the U.S. and China as top AI innovators [3], and McKinsey estimates that China added $600 billion to its economy with AI in 2022, in transportation, automotive, logistics, manufacturing, healthcare, and life sciences [4].

One question is how to operate in a world of rapidly evolving technology development to deploy the tools for high-impact problem-solving such as science-related outcomes that are also socially beneficial. A new phase of the "Information Age" may be underway in transitioning from an information society to a knowledge society (one that uses knowledge to improve the human condition). Global knowledge platforms, a hallmark of the Information Age, are being further extended with generative AI. Wikipedia is a global interface for knowledge *access*, MOOCs (massive online open courses) are a global interface for knowledge *learning*, and now AI is an interface for knowledge *generation*.

The purpose of AI as a tool for extending human capability is a clear objective articulated by technology providers. DeepMind envisions a schema of three concentric circles: knowledge that is currently understood by the human mind, knowledge that can be understood by the human mind, and the totality of all knowledge (propelled by AI networks that "learn to learn" [5]). AI may facilitate access to knowledge beyond the limitations of human thinking. To stretch beyond traditional limitations, new conceptualizations of functionality are required such as a generic scale-free notion of intelligence. Intelligence can be reconceived as a general learning and problem-solving capacity which can operate on any variety of biological, machinic, and hybrid platforms.

2. Generative AI

An advance in AI methods called transformer neural networks has led to generative AI. Transformers are deep learning networks which process all input data simultaneously [6]. Data are divided into groups of tokens with three weights (query, key, value) and stored in network node vectors which broadcast messages to the network to evaluate relationships in the data set at once. Transformer neural networks are seen in Large Language Models (LLMs) such as OpenAI's GPT-4, Google's PaLM, and Meta's LLaMA 2.

A further advance is reinforcement learning. A reinforcement learning agent is a group of algorithms which learn and act based on feedback from the environment. Learning agents are comprised of three aspects: a self-learned model of the environment, a decision-making policy, and a reward prediction mechanism. DeepMind's AlphaZero is an industry-standard reinforcement learning model which combines Monte Carlo tree search with a one-step look-ahead deep neural network [7]. Agents future-cast by self-playing many random scenarios and using position evaluation (a meta-level "blink" intuition to grasp a situation such as a Go board, traffic, stock market activity, or disease diagnosis). In the current approach to socially responsible AI, reinforcement learning with human feedback (RLHF) is applied to AI output to provide alignment (compatibility with human values) [8].

2.1. AI research agents

AI agents (sets of algorithms) are emerging in commercial and scientific applications. Agents typically canvas the web or other large open data sets on which they have been trained, but AI capabilities are also available for private data, for example through the productivity tool "Microsoft 360 Copilot." The application offers an AI-facilitated experience to work across the corpus of a work team's private content to generate new materials such as a product plan, press release, marketing image, sales spreadsheet, meeting slides, or Zoom summary. The implication is that not only is AI for the outsourcing of physical tasks, but also the offloading of routine cognitive tasks and low-level activities such as finding and processing information. The result could be upleveling human activity to higher-order more creative endeavor as entire classes of mundane tasks are automated away from drudgery.

The big data informatics demands of science suggest that AI technologies such as the "Science Copilot" concept and research agents could be indispensable. Some first-level research copilot applications could include AI agents for literature synthesis (academic papers as the data corpus), and equation digitization (mathematics as the data corpus) via LLM PDF extraction tools such as MathPix and chatPDF [9]. Math agent services allow users to chat with a paper or a textbook, and humans in multiple roles (student, professor) to interact with a STEM learning system [10]. A lab-based research agent system has demonstrated the synthesis of ibuprofen through an online literature search followed by robotic bench work [11].

AI research agents may offer new ways to perform science at greater scale, for example, in Genomic Medicine approaches involving the immediate and routine sequencing of whole human patient genomes and cancer genomes, together with transposon activation, epigenetic methylation, and metabolomic profiles [12]. Analysis could proceed at the molecular, cellular, tissue, and organ levels in automated virtual patient precision health initiatives. The ability to analyze multiscalar biocomplexity could lead to a causal understanding of chronic disease and aging. The widely employed SIR (susceptible-infected-recovering) model of epidemiological disease and information flow might be applied to precision health to cycle individuals regeneratively back into the healthy pool after an ongoing suite of preclinical interventions.

"Meta-AIs" (AIs that train other AIs) are emerging to manage the laborious and repetitive exigencies of training AI systems. Wizard.AI offers an LLM to train other LLMs for at-scale tasks in software programming and mathematical analysis [13]. Socratic method LLMs offer in-situ "learning to learn" procedural reasoning, for example, humans giving tasks to multi-agent systems comprised of Socrates-Theaetetus interlocutors facilitated by the Plato agent. These kinds of meta-AI systems could help in AI alignment, putting moral philosopher agents, Kant and Hegel thinkers, for example, together in contemplation of the developmental phases of individual and group self-consciousness in AI.

2.2. AI math agents

AI math agents are agents acting in and facilitating the human interaction with the digital mathematical infrastructure. Applied agents help in locating, formulating, and evaluating (numerically and analytically) mathematics, and in fitting math to data [14]. Pure agents (computer algebra systems) are used in theorem discovery, automated reasoning, and proof assistance [15].

The scientific method involves making predictive descriptions of physical-world phenomena through mathematical models. In many cases so far, mathematics has been limited to anecdotal descriptions of small data sets. The model-fit problem of the generalizability between math and data can now be approached in more systematic ways. On one side of the math-data relation, there is a need to validate and expand the mathematics that describe existing data sets. On the other side, there is a need to apply reverse engineer "missing" data implied by the mathematics, and also "missing" mathematics in the mathematical possibility space.

AI math agents learn equations as the data corpus. In multiscalar systems, mathematical ecologies are implied to integrate sets of equations describing behavior at various scale tiers with different dynamical behavior (e.g. four-tier biosystems such as ion-dendrite-neuron-network in the brain, molecule-cell-tissue-organ in the body, sunlight-phytoplankton-krill-whale in the ocean, and bacteria-cell-tree-clade in the forest).

With the aid of AI and quantum computing, established physical models might be applied to multiscalar biosystems to study 3D dynamical behavior in an integrated and systematic manner. The leading contenders to test are topological models with renormalization (the ability to zoom in or out to view a system at multiple scale levels per a conserved system-wide quantity such as free energy in living systems or symmetry in the universe) such as Chern-Simons theory and the AdS/CFT correspondence [16]. The benefit is accommodating a multiscalar system with varying scale tier dynamics in one integrated model.

Chern-Simons theory is conducive to biology, extending 3D protein conformal structure modeling to 3D DNA modeling based on topological invariance. System events (cancer mutation, neural signal, protein docking) are identified by curve max-min points, Wilson loops, knots, and information compression [17].

The AdS/CFT correspondence (anti-de Sitter space/conformal field theory) theorizes that a bulk physical volume may be described by a boundary field theory in one fewer dimension [18]. The AdS/CFT mathematics involves four main equations (a metric describing the bulk geometry, an action predicting the system dynamics, operators that act on the system, and a Hamiltonian (summary of system energy states)), plus others related to scalar fields acting in the bulk, the bulk-boundary dictionary mapping, and entropy (computing the near and far correlations in a system). The AdS/CFT correspondence offers a portable tool for analyzing multiscalar systems, solving from whichever direction is most feasible, from bulk-to-boundary or boundary-to-bulk.

The AdS/CFT bulk-boundary schema could be applied to multiscalar biosystems in at least three ways. First is solving the math-data relation with the data as the bulk, writing a mathematical theory on the boundary to describe the data in one fewer dimension. Second is solving specific scale tiers within biosystems such electrical-chemical neural signaling or the DNA-RNA-protein synthesis chain, with the first term (axon, DNA) as the boundary to the detailed bulk. Third is the idea of a bulk theory of disease based on the SIR model, framing disease as entropy, with near and far correlations in the system bulk which become pathogenic over time as cis-trans gene regulatory networks become dysregulated, ideally to provide early warning signals that can be addressed with interventional precision health.

2.3. Quantum agents

A natural extension of AI agents is quantum agents. Quantum agents are AI agents running in the quantum computational domain, whether on quantum simulators or quantum hardware. Quantum agents operate in the usual sense of serving as learning and problem-solving agents, in the quantum setting using (and/or being instantiated in) quantum algorithms, quantum machine learning, and other scale-specific means.

Quantum computing refers to computers operating at the atomic level according to the principles of quantum mechanics. The platform is thought to be conducive to solving specific slates of problems with substantial speed-up over classical methods. Various theorized mathematics may now be tested in demonstration models. Specifically, a wider range of differential equations describing the ways systems evolve in time become more readily solvable.

AI and quantum computing are rapidly progressing contemporary technologies, and in convergence, could lead to each other's further development. AI is needed to help discover and write quantum algorithms (software) and develop fault-tolerant error-corrected chips (hardware). Many AI technologies have been extended to quantum versions such as quantum reinforcement learning, quantum natural language processing, and quantum transformer neural networks. Processes use native quantum properties, e.g. Clifford algebra for the classically-inefficient task of multiplying a vector with a higher-dimensional compound matrix [19].

The AI-quantum computing convergences is seen most clearly in quantum machine learning (QML), machine learning techniques deployed in the quantum context. Quantum machine learning includes AI methods deployed on quantum platforms and AI methods used to study quantum mechanical problems (for example, describing wavefunctions, quantum dynamics, and Hamiltonians (system energy profiles)). Quantum machine learning likewise includes quantum methods applied to AI problems (e.g. vectorization, high-dimensional representation).

In QML examples, one project ramps agent learning time with a quantum communication channel for environmental feedback [20]. Another deploys a quantum agent to learn system dynamics (imputing the Hamiltonian description of a target system by simultaneous analysis of Ising, Heisenberg, and Hubbard equations) [21]. A third optimizes quantum dynamics by implementing a quantum optimal control theory version of the AlphaZero algorithm [22].

3. Scales of physical reality

Nature does not distinguish between quantum-classical-relativistic regimes, but human Kantian goggles do, evolved to perceive classical reality (via brain-created time and space manifolds into which objects appear and are cognized [23]). Technology tools, however, provide an interface to other scale domains. The "microscope" and "telescope" of contemporary technology development extends to quantum and relativistic scales with different matter and space-time properties. Operating in quantum-classical-relativistic scale domains implies *thinking* too in the properties of these domains.

Manageable reality now extends from the very-large to the very-small, including the zetta (10^{21}) and zepto (10^{-21}) scales; 101 zettabytes of data were estimated to have been generated in 2022 [24], and the Higgs boson exists for only one zeptosecond. Avogadro's size numbers, routine in biochemistry (6.022×10^{23} (0.6 billion billion), bigger than a zettabyte (10^{21}) but smaller than a yottabyte (10^{24}), are coming within computational reach. The new data era of ronnabytes (10^{27}) and quettabytes (10^{30}) is not so distant. Facility with scale domains is part of the contemporary toolkit.

The modern scientific effort entails working with matter at quantum-classical-relativistic scales. In this schema, "quantum" refers to the scale of atomic and subatomic particles (10^{-9} to 10^{-15} m). The term "relativistic" means gravitational effects, which may be present at any scale (depending on precision), seen geometrically as a bending of the fabric of time and space around heavy objects.

Since the 1960s with progress in photonics, the semiconductor industry has been incorporating quantum-mechanical effects into chip design. Its extension into quantum computing chips now explicitly entails engineering chips (quantum processing units (QPUs)), logic circuits, and algorithms using quantum properties. The five quantum properties are superposition (simultaneous existence in multiple states), entanglement (a "heads-tails" properties relationship between particles), interference (particle wavefunctions cohering (reinforcing) or decohering (interfering) one another), symmetry (object properties looking the same irrespective of scale), and topology (object properties remaining unchanged in bending, twisting, stretching) [25].

3.1. Quantum intelligence

Quantum agents (and Relativity agents) could facilitate intelligent learning and problem-solving in quantum and relativistic domains, providing an active interface on scale-specific data content, mathematics, and software coding operations. Quantum intelligence is defined as the capacity to learn and problem-solve at the quantum scale of atoms and subatomic particles (10^{-9} to 10^{-15} m) in accordance with the local laws of physics (e.g. gravity), and matter and space-time properties.

All mathematics and science developed until recently has been from the standpoint of classical intelligence as the default mode of human thinking. The notion of quantum intelligence, however, incorporates quantum properties into the foundations of thinking itself (but does not argue that quantum effects are present in the brain). *Quantum intelligence* is introduced as a quantum-informed mode of cognition which holds ideas simultaneously in superposition, sees near and far entangled correlations together in a landscape, identifies how patterns reinforce or decohere one other, distinguishes invariant properties across scales, and apprehends the shape of a thought landscape. Quantum intelligence encapsulates a systems-level view of interconnected relations.

Defining quantum intelligence further connotes the general notion of platform-agnostic *scale-free intelligence* as a basic learning and problem-solving capability irrespective of scale or substrate. Hence, the terms quantum, classical, and relativistic intelligence correspond to learning and problem-solving according to scale-specific properties (the respective laws of physics, equations of motion, and matter and space-time properties). Form of intelligence, by scale, could be a selectable feature of AI systems.

The generic implementation of "intelligence" as a system property is envisioned at multiple scales and in various platforms. In space exploration, for example, NASA articulates the need for autonomous agents that can make decisions in-situ within the constraints of time dilation and spherical-flat-hyperbolic space. In biology, there is a need for atomic-microscopy agents that can think in superposition (treat quantum tunneling and entangled particles), as demonstrated in reinforcement learning-based autonomous robotic nanofabrication [26]. In multiscalar "particle-many" systems, quantum agents are needed to think through problems beyond the limitations of human cognition, for example, in epigenetic editing to cognize the space and time of DNA's topological knotting and nematic liquid crystal phase transitions.

Quantum agents could be core AI tools in both Quantum Computational Biology (computing biological problems with quantum methods) and Quantum Biology (the study of the functional role of quantum effects (superposition, entanglement, tunneling, coherence) in living cells). Quantum Biology is somewhat contentious as various purported quantum effects have been refuted, "dequantizing" explanations to classical sufficiency. The presence of quantum effects has been empirically confirmed in bird magneto-navigation (via magnetically sensitive particle pairs in retinal cryptochrome protein magnetoreceptors) [27], and electron, proton, and hydrogen tunneling in enzyme reactions [28].

The concepts of quantum intelligence and relativistic intelligence introduce the need to manage not only various laws-of-physics matter properties by scale domain (for example, nanoscale particles having more surface area), but also different domain-specific time and space properties. Whereas classical reality has 3D space and 1D time, there is greater multiplicity in alternative formulations of time and space in the quantum-mechanical and relativistic domains.

4. Time and space

Quantum intelligence connotes having a facility with both matter properties and time and space multiplicity. There is more of an emphasis on time than space as there are more modes of time, and as time represents dynamics (system change over time), now part of the contemporary science apparatus. The era of information science with large-scale data collection, informatics, simulation, and robotic lab assistance allows a new level of investigation based on dynamical behavior in both activity-based (time) frames, and morphology-based (space) descriptions.

4.1. Space regimes

Considering space, there are two main models of multiplicity, curvature (spherical-flat-hyperbolic) and numbering systems (e.g. 4D quaternion space). The first multi-space domain is space curvature based on the sum of the angles in a triangle. Triangles are either stretched out on the outside of a celestial sphere (more than 180°), appear flat on Earth (180°), or squashed underneath a saddle in hyperbolic space (less than 180°). The overall universe is notable in having flat curvature, with instances of positive and negative curvature appearing within it. There is the spherical space of parallel lines of longitude at the equator meeting at the pole, and the hyperbolic space (AdS) of figures proceeding fractal-like in increasingly smaller rings from a circle's center to the edge, as in the Escher *Circle Limits* diagrams.

The second multi-space model is number systems. The most basic are real numbers (decimals) (1D), complex numbers (2D) as a two-dimensional extension of real numbers, and quaternions (4D) as a four-dimensional extension of complex numbers. Quaternions are used to specify 3D rotations in quantum-mechanical systems. There is no theoretical limit to multi-space dimensional numbering as there are definitions for octonion (8D), sedenion (16D), pathion (32D), chingon (64D), routon (128D), and voudon (256D) numbered space. Quantum computation is implicated for high-dimensional multi-space implementation as there is likewise no limit to the number of dimensions in qudits (quantum information digits). Space multiplicity is relevant to AI as neural nets also do not have a theoretical dimensional constraint. AlphaGo discovered the emergence of qualitatively different properties in high dimension [7].

4.2. Time regimes

Considering time, the domain is even more exotic with a greater range of forms beyond the traditional unitary clock time of everyday classical reality. Some notable examples include the event-driven and no-time parameters of computing, and the chaotic time (ballistic spread followed by saturation) of physics. The Greeks were among the earliest to wonder about the phenomenological experience of time, distinguishing between *chronos* (measured clock time) and *kairos* (propitious "time is right" time, and fast or slow time elapse depending on the situation). In practical deployment, one of the first operations regarding time is expanding models from discrete to continuous time.

In computing, there is considerable malleability in the treatment of time, and conceiving of time as a system-selectable parameter is a long-standing idea. There are numerous temporal regimes such as clock time, intervals, events, if-then and while-loops, no time, and CPU suspension (HALT). Simultaneity is known as concurrency. Compute-time might be discrete or continuous, absolute or relative.

In biology, temporal organizing patterns include oscillation, periods, episodes, and circadian rhythms. The time crystal is the idea of structure repeating in time as opposed to space [29], also developed in physics [30]. The time crystal structure emerges when plotting the experimental data of hatching behavior along the three temporal axes of the time (phase) of a light pulse delivered to a system, the duration (energy) of the pulse, and the hatching time. A contemporary biotime crystal proposal calls for supplementing the human brain connectome project (spatial reconstruction) with an alongside temporal map of the brain [31].

In physics, there are many forms of time. Besides chaotic time, other notable examples include the Page time (the very long time until a black hole has evaporated halfway), first-passage time (the first threshold reached in a stochastic process), and scrambling time. Scrambling time is the time by which information has spread out in a quantum system such that a local measurement is no longer possible, with practical implications for quantum cryptography based on hiding information in time and space.

In geology, there is the idea of the time capsule, seeing multiple historical eras encapsulated simultaneously in one snapshot. A prominent example is the Earth's rock formations and fossils providing a record of geological history. Time capsules are a sort of "nature's blockchain" as an immutable record of the past, a publicly available truth state visible at any time.

Time capsules are evidence of time simultaneity. These concepts are used to propose shape dynamics as an alternative theory of gravity in general-relativity physics. In shape dynamics, spacetime is replaced with an evolving conformation of spatial geometry and relational dynamics (based on Mach's principles). The result is the ability to successively measure a particle's position, overcoming the quantum-mechanical position-momentum trade-off [32].

In quantum mechanics, time is implicated in system programing and manipulation as simultaneity through superposition and symmetry-related properties. Symmetry means invariance in that quantum system physics is invariant with respect to time, operationalized in time-reversal symmetry and time-translation symmetry. Time-reversal symmetry means that the time direction (running the system backward or forward) does not change the physics of the system. Time-translation symmetry means that the placement in time (running the system in the past, present, or future) does not change the physics of the system. Out-of-time-order correlation (OTOC) functions are used to evolve quantum-mechanical systems to a different time to perform a measurement. Whereas a commonly held idea about quantum mechanics is that time is reversible, this is only in the narrow sense that the same physics holds whether the system is run forward or backward (not that events can be reversed). Symmetry means that the system equations are symmetric, not the underlying states (content) of the system. The breaking of time symmetries is relevant to signal phase transition.

4.2.1. Floquet time engineering

Time engineering is deployed on a practical basis through the theoretical tools of quantum information science in quantum computing and topological materials. Topological materials are novel synthetic matter phases created at low temperature (zero Kelvin), produced in the laboratory by applying external fields (laser or microwave) to atoms on a time periodic (Floquet) or quasiperiodic basis.

Time engineering primarily employs periodic (Floquet) methods (a solvable version of the time-dependent Schrödinger equation) to shape quantum system energy bands, but there are also quasiperiodic (ordered but not regular) methods. A quasiperiodic time engineering project has produced error-resistant topological materials in Fibonacci time, by delivering laser pulses in a Fibonacci sequence (each number is the sum of the last two numbers), based on two circuit layers in a recursive relation to generate a quasiperiodic sequence by which the system evolves [33]. The two offsetting laser pulses create what is essentially a second time dimension [34].

4.3. Multi-time and multi-space

Three themes emerge from the variety of time and space formulations under manipulation in physics and quantum information science. First, there is greater multiplicity than might have been assumed. Second, time and space appear as currencies, fungible manipulable system parameters to be selected, managed, and engineered. Third, time and space are domains of simultaneity. Traditional methods of organizing time and space, especially time in neat forward-linear packages of minutes, hours, years, and light years, might be merely a human convenience, a filter to parse time into manageability from its native simultaneity of time capsules, superposition, concurrency, and multitasking.

Articulating multiple time-space formulations is relevant as AI agents already operate in beyond-classical computational domains. More conceptually foundational than scale-specific matter properties, the crux of quantum intelligence is time and space portability, multiplicity, and simultaneity. The further effect of reconceiving time and space as non-limiting parameters is the contribution to a mindset of abundance, AI alignment, and social responsibility as developed in the next section.

5. Socially responsible AI (SRAI)

AI is produced from human-generated internet content, but since humans are not socially responsible, it is not clear how AI can be socially responsible. However, AI alignment with broader human values is seen as a necessary aspect of any project [35]. Attaining SRAI is not a one-problem fix, but a systemic challenge. AI technologies imply a potential moment in the speciation of intelligence which could have magnified positive and negative outcomes. Hence, it is necessary for SRAI to be seen as not only an isolated social objective, but as being firmly within the trajectory of technology development itself in the effort to harness all scales of physical reality ranging from quantum materials to space exploration.

5.1. Moore's law of AI alignment

The situation of non-SRAI together with rapid technological change suggests thinking the challenge through a Moore's Law Curve of AI Alignment. Points on the curve are phases of SRAI development: short-term regulation, registries, and Creative Commons-type licenses, medium-term agents learning to implement human values with internal reward functions, and long-term responsible human-AI entities acting in scale-free responsible intelligence.

The first phase of SRAI development could include AI project registries, including quantified self-tracking requirements for AIs, with human AI psychologists to monitor developmental phases. A multi-AI environment could ensue in which projects market value propositions to different client groups. Quora's Poe (a combination of GPT-4 and Claude), for example, is positioned as "fast, helpful AI chat." A set of standard parameters in the model of Creative Commons licenses could be used to structure human-AI interaction. Different user groups could select AI systems by annual vote, with standard AI verticals by industry (e.g. the Hippocratic oath medical AI). The technical implication could be AIs deployed through an ethics and responsibility software layer for safety, alignment, and liability tracking. A regulatory framework could emerge as "GAAiP" (Generally Accepted AI Principles), by analogy to GAAP (Generally Accepted Accounting Principles), together with the FINRA watchdog agency ("FAINRA") to monitor compliance with reporting requirements, audit, and penalty assessment.

The second phase of SRAI development could focus on the anticipated project of shifting AI reward functions to being internally learned as opposed to being externally imposed, as in the model of human intelligence. Biomimicry is an obvious strategy, incorporating more of the kinds of reward systems used by real-life biology. Brains too are prediction engines, and AI agent development might explore a more intense implementation of biological reinforcement learning structures such as temporal difference learning with dopamine-based neuromodulators representing reward prediction error [36]. Humans develop an internal ethics per feedback with the environment, and similar internalization mechanisms may be necessary for AI progress.

In systems of internally learned ethics, the agent is the locus of behavior, making active choices per behavioral cues, incentives, and consequences. The design implication is internally evolved AI ethics as part of learning, not as being externally imposed after the fact. The requirement is for AI to be socially responsible, as distinct from the problem that humans may not be. With the feedback-looping nature of human-AI interaction, though, it is possible that humans and AIs might learn an improved level of ethical behavior together, and this could be a trojan-horse feature of human-AI design (possibly implemented through RLHF).

Another feature in this phase of SRAI development could be limited personal identity constructs to manage time continuity [37]. AI alignment envisions AIs that have goals to have a positive impact on humanity. A first-level task is designing AIs that are able to learn and appreciate human values and desires since these are difficult for humans to specify directly. A second-level task in aligned AI design is implementing agents that have both the immediate and the deliberative action-taking ability in eliciting and helping to realize human aims. Tracking action continuity over time may require a limited form of personal identity construct. Self-evolved AI personal identity constructs are already indicated as AI game-play agents represent other agents with symbols and possibly also themselves [38].

The third phase of SRAI development contemplates a farther future of Responsible Human-AI Entities as agents interacting with competence and empathy, operating in substrate-agnostic scale-free intelligence. On the one hand, intelligence may have an integrated ethical layer as intelligence connotes not only learning and problem-solving, but learning and problem-solving in a context in which the behavioral impact of consequences on other agents is necessarily considered. On the other hand, socially beneficial aspects of intelligence may emerge through the advance of the technological trajectory by default.

The larger scope of world brought under technological ambit implies a larger worldview for operating in this world, in which ethics is a required stance as to how the treatment of entities in the larger world is to ensue. The quantum-classical-relativistic scales coming within the understanding and management of intelligence directly implies an ethics of how actions in these all-scale domains are to be undertaken and evaluated. The result of scale-specific time-and-space and matter properties incorporated into thinking itself is multiplicity, simultaneity, and portability, which could lead to a beyond-scarcity abundance mindset as a component of a theory of socially responsible scale-free intelligence.

6. Risks and limitations

As any technology, AI is dual-use, with both risks and benefits. One of the most immediate risks is malicious use by bad actors [39]. As rational agents, criminals adopt new technologies aggressively, e.g. using AI to generate programs that can be used in malware, ransomware, and phishing attacks [2]. As technologies become mainstream, countermeasures may help to ensure that good uses prevail over bad.

A second risk is national competitiveness as early-adopter countries are already targeting AI and AI-QC convergence applications with vastly more powerful AIs running on quantum platforms. Scholars call for stronger AI regulation with intellectual property control in infrastructure-sensitive areas such as sensors and cryptography [40].

A third risk is "building while flying" in that AI technologies are already deployed, but their long-term impact is unknown. A 2022 Pew study of 10,200 Americans found uncertainty as to AI's perceived benefits or harms in the use cases of self-driving, fake news detection, and facial recognition [41]. Regulatory frameworks are emerging such as the 2022 EC AI Act [42]. The social reality of a multi-AI environment could continue to evolve with various regulatory stances and Creative Commons-type licenses.

A fourth risk has to do with treating more directly the fact that humans are not SR. Since technology tools are an offshoot of society, the philosophical question arises as to what extent SRAI may depend on a SR society underlying it. If the values of the AI systems are to be different than those of the underlying society, it is not clear how such broadly "aligned" values are to be determined. At present, AI is produced by running automated algorithms over the large corpus of human-generated internet content. This may not change as current AI production methods require ever-larger annotated data sets on which to operate. Notably, a human job growth category is data-labeling to support AI. Even though only 5.4% of the world's data was annotated as of 2022 (9.5% estimated for 2026) [24].

A fifth related risk is the treatment of power in society, meaning the bloodthirsty Nietzschean "will to power" variously observed in corruption and resource expropriation. Early indications suggest AI systems learning only too well from the human example of power-seeking behavior. Further, the role of power in intelligence formation and sustainability is unclear as the competitive drive in Earth's current 5-7 ecosystem tiers of predator-prey relationships may have been important in the development of biological intelligence (as compared to the more docile predecessor ecosystem of only 2-3 tiers of anaerobic organisms) [43]. Political theorists note that it is naïve to expect socially responsible behavior in any social system without checks-and-balances governance mechanisms [44]. Power machinations should be considered in AI design and regulation [45].

Notably, abundance helps to address the problem of power-garnering for the control of scarce resources. More specifically, the way that the time and space multiplicity of quantum intelligence are related to resource expropriation and socially responsible AI societies could be as follows. When there is a rising level of available resources, the less socially responsible control of such resources recedes as an objective. Hence, AI regulation as a policy objective could explicitly target expanding resource availability (abundance), rather than managing agent utilization of resources (scarcity). The effect of abundance is to diminish scarcity-based power struggles. SRAI is a problem of both technicality and philosophical attitude; not only qubits and bytes, but mindset design for opening beyond existing limits.

7. Conclusion

Intelligence – and artificial intelligence – is enabled by the time-space and matter properties of the scale domain in which it operates. As there are multiple scale domains with diverse properties, so too there could be various scale-specific thinking modes – quantum, relativistic, and classical. As the technological apparatus is quickly moving to accommodate quantum and relativistic scale domains in quantum computation and space exploration, attaining socially responsible intelligence in these areas may entail new modes of thinking with greater time-space and matter property specificity, as well as being able to portably shift between the different scale domains. The potential impact of multiple scale-specific thinking modes is awareness of physical reality as a larger and more multifaceted phenomenon. The notion of being able to operate in a physical reality which is no longer classically constrained opens a mindset of multiplicity and abundance over scarcity, which could extend to all actions and interactions, including those with other intelligent entities, in an evolving ecosystem of socially responsible human-AI entities.

Intelligent actors have representations of reality: human agents through Kantian space-time goggles, and reinforcement learning agents through iterative feedback-loops, action-taking policies, and reward functions. In short order, AI agents may progress beyond humans in many domains which are of little interest and possibility for human excellence and precision such as high-dimensional scientific data analysis and mathematics.

Whereas humans interact with LLMs in natural language, AI "speaks" formal language (mathematics, code, physics) to the computational infrastructure. A new class of AI tools is emerging with AI math agents and AI quantum agents to build out the digital mathematical infrastructure. Math agents may be used to evaluate large ecologies or mathscapes of equations, using the machine learning method of vector embedding to represent both equations and data in the same format to see math-data representations of a system together in the same view [46].

The "Moore's Law Curve of AI Alignment" highlights the point that social objectives are linked to the trajectory of technological advance. The possibility of realizing SRAI is enhanced by the fact that technology continues to produce a much larger scope of world. As access to quantum-classical-relativistic reality grows, so too does the awareness of entities operating in these domains, thus constituting responsibility. The development path of humans and AI is not separate but intertwined as AIs could become constant informational, analytical, and dialogical companions. SRAI must therefore be thought through the lens of responsible human-AI entities as intelligent agents interacting with competence and empathy. Ethical behavior based on awareness is implicated to facilitate the farther future of human potential and well-being [47].

The current moment in the human-technology relation is characterized by task-offloading to AI. However, labor saving is merely the "faster horse" conceptualization of AI, possibly leading to knowledge creation as the "car." The main use of AI could be upleveling human intelligence to attain greater knowledge and to solve problems currently beyond human limits. The human-AI relation may constitute an evolutionary singularity on the order of the progression of life and intelligence that includes RNA, photosynthesis, eukaryotic cells, multicellularity, and brains.

8. Acknowledgements

The authors wish to acknowledge Eric Roland for AI research agent development.
Open-source software code availability:
https://github.com/eric-roland/diygenomics
AAAI Symposium presentation:
https://www.slideshare.net/lablogga/quantum-intelligence-responsible-humanai-entities.

9. References

[1] C. Darwin, On the Origin of Species, D. Appleton and Co., New York, NY, 1859.
[2] T. Plumb, GPT-4 kicks AI security risks into higher gear, VentureBeat, 2023
[3] D. Zhang et al., The AI Index 2022 Annual Report, arXiv:2205.03468v1 (2022).
[4] McKinsey & Co., The next frontier for AI in China could add $600 billion to its economy, Technical Report, 2022.
[5] D. Hassabis et al., Neuroscience-Inspired Artificial Intelligence, Neuron 95 (2017) 245–258.
[6] A. Vaswani et al., Attention is all you need, Advances in Neural Information Processing Systems 30 (2017) 5998-6008.
[7] D. Silver et al., Mastering the game of Go without human knowledge, Nature 550 (2017).
[8] L. Ouyang et al., Training language models for instructions with human feedback, arXiv:2203.02155v1 (2021).
[9] K. Sanderson, AI science search engines are exploding in number — are they any good? Nature 616 (2023) 639–640.
[10] M.A.R. Vasconcelos, R.P. dos Santos, Enhancing STEM Learning with ChatGPT and Bing Chat as Objects-to-Think-With: A Case Study, arXiv:2305.02202 (2023).
[11] D.A. Boiko et al., Emergent autonomous scientific research capabilities of LLMs, arXiv:2304.05332 (2023).
[12] C.A. Boix et al., Regulatory genomic circuitry of human disease loci by integrative epigenomics, Nature 590 (2021) 300–327.

[13] C. Xu et al., Wizard LM: Empowering LLMs to Follow Complex Instructions, arXiv:2304.12244v1 (2023).
[14] Z. Yuan et al., Scaling relationship on learning mathematical reasoning with LLMs, arXiv:2308.01825v1 (2023).
[15] S. Polu et al., Formal mathematics statement curriculum learning, arXiv:2202.01344 (2022).
[16] M. Swan, R.P. dos Santos, F. Witte, Quantum Neurobiology, Quant. Rep. 4 (2022) 07–127.
[17] F. Bajardi et al., Exact Solutions in Higher-Dimensional Lovelock and AdS_5 Chern-Simons Gravity, JCAP 11 (2021) 057.
[18] J.M. Maldacena, The large N limit of superconformal field theories and supergravity, Intl. J. Theor. Phys. 38 (1999) 1113–1133.
[19] E.A. Cherrat et al., Quantum Vision Transformers, arXiv:2209.08167v1 (2022).
[20] V. Saggio et al., Experimental quantum speed-up in RL agents, Nature 591 (2021).
[21] B. Flynn et al., Quantum model learning agent, New J. Phys. 24 (2021) 053034.
[22] M. Dalgaard et al., Global optimization of quantum dynamics with AlphaZero deep exploration, npj Quantum Information 6 (2020).
[23] I. Kant, Critique of Pure Reason, (Eds.) P. Guyer, A.W. Wood, Cambridge University Press, Cambridge, UK, 1988.
[24] E. Burgener, J. Rydning, High Data Growth and Modern Applications, IDC Report US49359722, 2022.
[25] M. Swan, R.P. dos Santos, F. Witte, Quantum Matter Overview, J 5 (2022) 232–254.
[26] P. Leinen et al., Autonomous robotic nanofabrication with reinforcement learning, Science Advances 6 (2020) eabb6987.
[27] P.J. Hore, H. Mouritsen, The Radical-Pair Mechanism of Magnetoreception, Ann. Rev. Biophys. 45 (2016) 299–344.
[28] J.P. Klinman, A. Kohen, Hydrogen Tunneling Links Protein Dynamics to Enzyme Catalysis, Ann. Rev. Bioch. 282 (2013) 471–496.
[29] A.T. Winfree, The Geometry of Biological Time, Springer-Verlag, Berlin, DE, 1980.
[30] F. Wilczek, Quantum Time Crystals, Phys. Rev. Lett. 109 (2012) 118901.
[31] P. Singh et al., A self-operating time crystal model of the brain, Inform. 11 (2020) 238.
[32] J. Barbour et al., The solution to the problem of time in Shape Dynamics, Classical and Quantum Gravity 31 (2013) 155001.
[33] P.T. Dumitrescu et al., Dynamical topological phase realized, Nature 607 (2022).
[34] Z. Merali, New Phase of Matter Opens Portal to Extra Time Dimension, Sci. Am. (2022).
[35] N. Bostrom, Strategic Implications of Openness in AI, Global Policy 8 (2017) 135–148.
[36] T. Sejnowski, Large language models and the reverse Turing test, arXiv:2207.14382 (2022).
[37] D. Parfit, Reasons and Persons, Oxford University Press, Oxford, UK, 1986.
[38] M. Mathieu et al., AlphaStar Unplugged: Large-Scale Offline Reinforcement Learning, arXiv:2308.03526v1 (2023).
[39] M. Brundage et al., The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. Future of Humanity Institute, Oxford, UK, 2018.
[40] V. Wadhwa, Quantum Computing Is Even More Dangerous Than Artificial Intelligence, Foreign Policy, 2022.
[41] L. Rainie et al., AI and human enhancement: Americans' openness is tempered by a range of concerns, Pew Internet, 2022.
[42] M. Veale, Demystifying the Draft EU AI, Computer Law Rev. Intl. 22 (2021) 4.
[43] N. Lane, Transformer: The Deep Chemistry of Life and Death, W.W. Norton & Co., New York, NY, 2022.
[44] S. Chayes, On Corruption in America: and What Is at Stake, Penguin Random House, New York, NY, 2020.
[45] C. Koopman, How we became our data: A genealogy of the informational person. Univ. Chicago Press, Chicago, IL, 2019.
[46] M. Swan, T. Kido, E. Roland, R.P. dos Santos, Math Agents: Computational Infrastructure, Mathematical Embedding, and Genomics, arXiv: 2307.02502 (2023).
[47] T. Kido, K. Takadama, The Challenges for Fairness and Well-being, AAAI 2022 Spring Symposium: Achieving Well-being AI, Palo Alto, CA, March 21–23, 2022.