Deep Science Publishing, 2026
https://doi.org/10.70593/978-93-7185-416-0_5

DeepScience
Open Access Books

Chapter 5: Principles of Agentic AI and Autonomous Decision-Making Systems

5.1. Introduction

A systems-oriented approach to AI Ethics coupled with a risk-based framework can inform the development of Agentic AI and Autonomous Decision-Making Systems capable of behaving in an ethical manner without the need for Manual Ethical Control. Artificial Intelligence (AI) an autonomous decision-making systems can significantly benefit humanity, as demonstrated by their deployment in a range of critical sectors including healthcare, energy, education, scientific research, and safety-critical activities. However, the development of Agentic AI with the capacity to independently achieve goals in a manner similar to that of humans remains a subject of active ongoing research.

The decision-making processes of these Agentic AIs can differ from traditional statistical decision-making processes, which are able to support Decision Automation and Guidance without the need for Manual Ethical Control. In contrast, Ethical Control Mechanisms deal with an entirely different aspect of AI safety. These systems have the capacity of Ethical Operation that remains incomparable to current AI and autonomous decision-making systems.

5.1.1. Executive Summary

Agentic AI refers to nondeterministic Autonomous Decision-Making systems comparable to humans or other sentient beings. It's defined as their capability for autonomous, nondeterministic behavior that entails the practice of writ large in the real-world environment. Such systems will act as agents running “small world” hypotheses of the decision-making problem for connecting the decision formulation and world modeling threads. The former thread that formulates the goal is the use of information about previous similar environments to a similar kind of model of the world; experimentation to reduce this uncertainty; observation of the actual world; the

formulation and resolution of the use of formulation of goal-based conjectures; or the operation of some set of utility functions over plans involving sequences of action and similar types of. The evaluation thread connects with the use of non-ideal but good enough heuristics that require verification and validation with respect to the environment.

Agentic AI must possess an architecture based on the neuromorphologic principles that underlie general intelligence that are be privy to direct use for hyper-realistic actual training of a sample concerned sub-portion of an Agented Domain. The Capability of Agentic AI must be assessed against international standards at each level of increasing autonomy or Agentic Level. The finessed set of benchmarks for Autonomy needs to assess with a suite of MCoH.

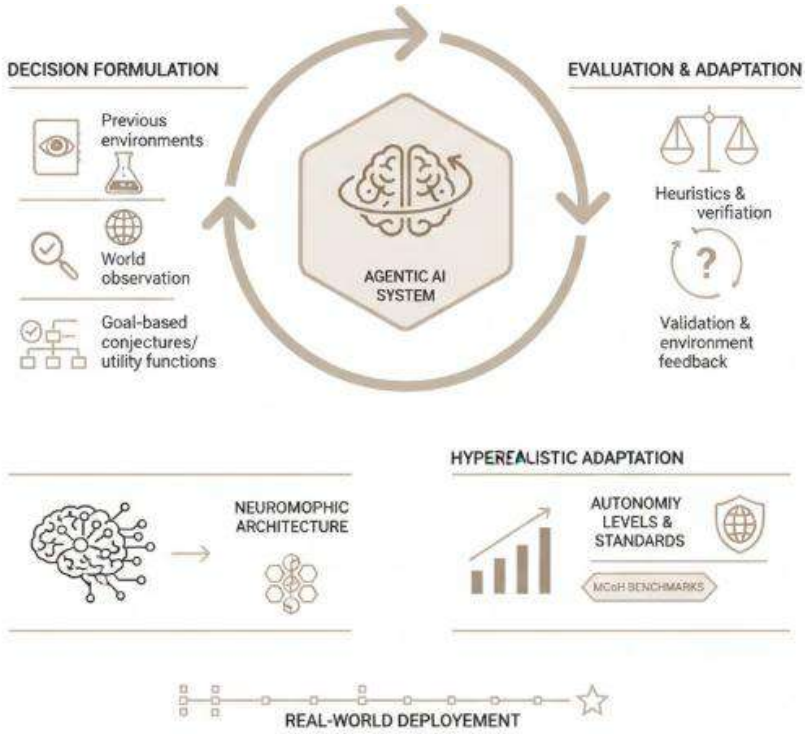

Fig 5.1: Neuromorphic Agentic AI: A Framework for Nondeterministic Autonomous Decision-Making and Multi-Level Autonomy Benchmarking

5.2. Foundational Concepts

The principles underlying Agentic AI and Autonomous Decision-Making Systems (ADM-Sys) advance to a higher academic and professional level as foundational concepts and arguments are more distinctly defined and elaborated. Concepts directly

derived from Clarks' op-ed, specifically Agentic AI and ADM, are offered as starting points for the development of an objective, evidence-based outline of principles, arguments, and implications. Agentic AI denotes a capability or class of systems to autonomously initiate, plan, and act without the necessity of continuous human supervision or constrained environment. The concept thus encompasses systems considered to have achieved Artificial General Intelligence (AGI) or a significant degree of intelligence along identifiable lines. ADM-Sys is a term that encompasses systems without the intelligence characteristic of AGI but that possess a facility enabling independent decision-making and actions for protracted periods.

The introduction of the concept of ADM-Sys acknowledges existing intelligent capabilities that facilitate independent decision-making and initiation of action yet fall short of human-level intelligence. The added clarity in distinguishing Agentic AI as a capability or class of systems and ADM as its objective enables specific characterization of ADM-Sys, an increasingly important and ethically problematic area of development across a range of human and robotic systems. Responsiveness of Agentic AI and ADM-Sys to the user's intention, approval, and requirements now occupy an evident role, and further discussion of concepts and the decision-making framework from which these considerations emerge is appropriate.

5.2.1. Agentic AI and Autonomy

Agentic AI, or Agentic Autonomous Decision-Making Systems, are defined as artificially intelligent systems that autonomously perceive their environments, make decisions, and take actions to achieve their objectives. The term agentic—having the capacity to act independently or to choose ones actions—stresses the systems capability to make policy decisions without requiring human involvement. Agentic AI encompasses a broad range of systems capable of higher-level cognition and decision-making, including the Classical Turing Test, the Google DeepMind GopherAI language model, and AlphaZero chess-playing algorithm.

The term autonomy describes the systems ability to self-govern by making independent choices within a complex, changing environment. The level of autonomy for which an AI system is designed is determined by the degree of deliberative assistance it requires from a human operator, driving development by an increasingly abstract decision-making layer. Such an autonomous system can still have an active human-in-the-loop (HIL) decision-making role, but the operation center moves from a supervisor role—monitoring and correcting—toward a plan-and-manage role, where human operators supply high-level directives and the agentic AI technology makes real-time decisions.

5.2.2. Decision-Making Frameworks

Decision-making frameworks should guide the monitoring, control, oversight, and mitigation of autonomous decision-making systems. Driverless cars illustrate why it is critical to apply formal frameworks to ensure that systems perform effectively in the simplest scenarios before moving on to more complex scenarios. The Detour Model, which specifically addresses the complexities of Agentic AIs, aids in evaluating whether an autonomous actor's proposed action is consistent with the norms, values, goals, and behavior of the human principal. Norm-based modeling of society (NMS) also contributes to the understanding of how complex human societies work and is an essential part of developing autonomy-aware language models that can reply consistently from a human society perspective. Norms provide information about a target system's goals, but norms alone are insufficient for proper orchestration, monitoring, and control of Agentic AIs.

The validation of decisions made by Agentic AIs, however, is a much more intricate matter. Beyond the implications for ethical agency, it has important consequences for accountability. Explanations for actions taken must be relayed to the relevant human stakeholders through an appropriate linguistic formulation or a combination of natural languages, graphical formats, and/or gestures. Empirical evidence suggests that comprehensibility is assisted by specificity and the attribution of goals and intentions to the agent being explained, making DSR a suitable building block. In doing so, the explanation should filter away the (presumably) less relevant aspects of the decision while highlighting the most important contributing causal elements, thus focusing on the most relevant part of the explanation. Researching the automated generation of explanations for the decisions of Agentic AIs remains an open area of work and will become increasingly critical as more capable Assistant-type systems are developed.

5.3. Ethical and Legal Considerations

The ethics and law of Agentic AI are tailored to decision-making systems. As autonomy increases, agency appears. Legal responsibility for agentic AI (and agents more widely) is not assigned to the AI, but to the organization, product owner, or legal person that acts and takes risks. Within a company, systems receive audits, overseers, and other checks. The law holds people accountable; within safe margins, people delegate responsibility.

Three criteria increasingly apply as agentic AI becomes autonomous and enters fields (such as health care or military operations) requiring a human controller: accountability and responsibility, transparency and explainability, and elucidation of safety, security, and other requirements.

Accountability and responsibility are ultimately legal and ethical concepts. When agentic AI reaches a sufficient level of autonomy and its application domain is associated with risk, action cannot be left unchecked. Internal testing, validation, and verification are insufficient. AI products in high-risk categories undergo continuous monitoring by an independent operator during deployment or are subject to predefined ex-ante requirements. With agentic systems, accountability and responsibility ultimately depend on the human Authorized User and the end customer. The Authorized User is responsible for assigning or delegating the task to the AI, defining the requirements, and for initiating action or controlling possible mission failure.

5.3.1. Accountability and Responsibility

The use of Agentic AI systems has far-reaching ramifications. On one hand, they could offer significant benefits to both the public and private sectors. Moreover, the control and direction of such systems could be fully embedded in decision-making frameworks that allow for a clear assignment of accountability and responsibility. This would separate the ethical responsibilities of the AI system from those of the human designers, supervisors, and managers.

On the other hand, the deployment of Agentic AI systems could carry existential risks, especially if advanced systems become uncontrolled or uncontrollable, Battlefield AI systems used in large-scale armed conflict could engage in runaway processes, or powerful autonomous agents pursue misaligned goals. As with any innovative technology, there could be serious consequences arising from unjustified complacency or unjustified fears. Novel mitigation approaches, capable of tackling the problems at hand in novel ways, could also be applied to novel classes of risk. It may, for example, be possible to distribute high-stakes alignment work across many agents, laying the groundwork for broad-based and robust alignment of increasingly autonomous decision-making systems.

5.3.2. Transparency and Explainability

Accountability, responsibility, decision completion, decision failure, consensus all arise when human actors design, deploy, and use autonomous systems. Successful decisions—those satisfying best outcome over all consequential considerations, including non-simple conditions and complex non-payment rewards—benefit all, including developers, users, owners—all direct actors and final actors of whom agents must gain transparent assurance. In essence, autonomous systems present their decision consequences and rationale solely through human time-honoured admissible game arguments, reframing the original complexity constraints. Human and agent final actors

achieve shared decision consensus by both accepting the agents' ruling into their external adaptive worlds.

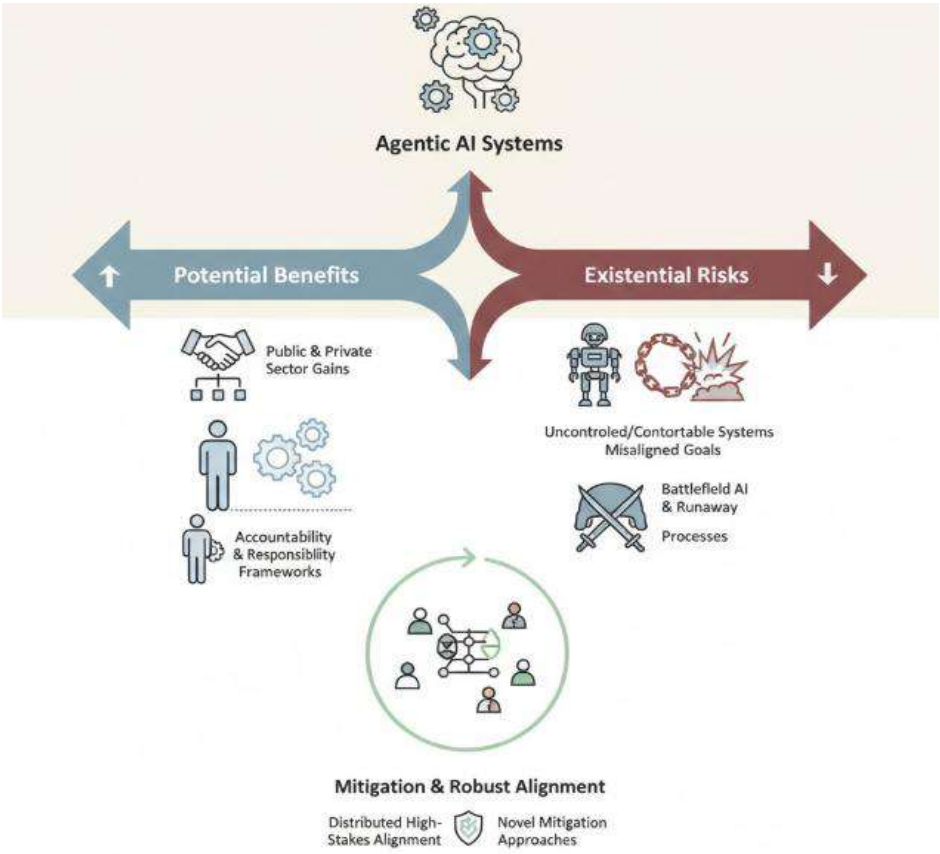

Fig 5.2: Navigating the Agentic Frontier: Governance Frameworks for Distributed Alignment, Accountability, and Existential Risk Mitigation in Autonomous Systems

Agentic systems that succeed in alleviating transparency risks operate through transparent decision rationale presentation complied with timely transparency expectations of final decision actors. Such transparent rationale opaqueness may potentially cause-transparency risks during the system evaluation phase. Agentic systems in final evaluation phases precede decision completeness through composite justification of final compliance. Such computational law ethnography serves final decision actors by ensuring absolute machine accuracy across current temporal expanse during decision completion. Machine Actor-based Parallel Path Independent Cryogenic Political Game® Protocol Momentum Theory® Circadian Cosmology brings assurance akin to scientific proof in final actor domain but must remain transparent at intermediate time-points.

Interventions of the external adaptive world seeking performance improvement designated on satisfaction of complex non-payment rewards should simply reveal performance-fractile experience, indeed performing a validation experiment on the agentic systems. Such design prevents embedded complexity or training time of final actors on non-reward experience or decisions. The transparent decision process can deviate from the expected ruling only when the ruling is perceived as a non-payment poor decision. Demand for alignment of non-payment or bad decision thus exogenously incorporates ethical imperative similar to that of agency into non-simple autonomous systems.

5.4. Architecture of Agentic Systems

Agentic autonomous systems must exhibit a minimal set of capabilities. First these systems require a perceptual capability that facilitates the acquisition of information external to the agent. From this information the agent constructs an internal model of the environment that is constantly updated over time. Such an internal model allows the agent to identify the provision of its internal goals and to formulate the means and actions required to progress toward these goals, steps which are driven by forecasting these actions when executed in the world. Finally, in the absence of explicit instructions concerning the satisfaction of those internal goals, the agent must exhibit the capacity to resolve these goals.

An agent's internal model of its environment and world updates is used both for world modelling and for goal identification and resolution. The capabilities required for such a goal resolution subsystem share much in common with the goal identification and planning subsystems typically found in cognitive architectures in that they are tasked with resolving and formulating means to satisfy the agent's internal goals. However, in the case of an agentic autonomous system, both of these functions have a specific characteristic, namely that they do not explicitly require an agentic structure. Rather than being a function of a social structure that can hold its principle responsible for action, such capabilities provide a means to reason about the agent's internal goals, or drive the agent when explicit instructions concerning those goals are not available.

5.4.1. Perception and World Modeling

An important aspect of agentic AI systems is the specification and implementation of the perception components that provide input to the decision-making modules. Perception entails defining the attributes of the environment associated with the chosen decision-making framework, identifying appropriate sensors for the targeted domain, and integrating data from these sensors into a consistent model of the environment.

The representation of the environment must include the results of the perception and world-modeling components, which combine sensory input with knowledge integrated from external sources, such as downloaded knowledge bases, simulation models, and expert feedback. The semantic and schematic information processing undertaken by these components reflects the capabilities of the autonomous actor, and it is important that the content of the environment representation be tailored to the function and competence of the particular agent.

Well-resourced systems, such as those tasked with synthetic media generation, may possess a detailed model of the real world, although often at significantly larger complexity than required for satisfactory performance. These models may incorporate significant orthogonal functionality, such as the production of natural language structured media. In contrast, other systems may employ much simpler, more domain-specific representations, for instance in a local security role, where the perception subsystem might provide a normality detector with inputs of recent history, exaggerated proximity cues, and a local blockchain-derived situation control tree.

5.4.2. Goal Formulation and Resolution

A goal is an internal description of a desired state of the world. Goal formulation is the sub-process of higher-level decision-making responsible for generating agentic driving objectives. It derives goals from a more abstract set of driver responsibilities. These driver responsibilities are agent-neutral descriptions of obligations, prohibitions, and permissions defining agent behaviour.

The most pressing developed goal formulation mechanism is that of Lang et al. [146]. The proposed mechanism abstracts a set of agent-neutral responsibilities – expressed using the recent concept of Ideal Accountability Languages – to a set of logical state constraints over the environment. The desired satisfying states of the environment are then generated using classical planning technology and mid-level plan generation decision-making. It is currently uncertain whether or not this type of goal formulation can be integrated well with agentic reasoning in Hybrid Decision-Making Architectures.

5.5. Evaluation and Validation

Evaluation metrics for autonomous decision-making systems must encompass three aspects. First, the evaluation should use a standardized set of decision-making models that reflect the current spectrum of capabilities being developed within autonomous decision-making systems. For simpler systems, a until-grid of benchmark problems solved under various uncertainty levels may suffice, although richer benchmarks also

exist for the critical task of navigation. More complex systems, like Open-World and General-Purpose agents, require progressively intricate decision-making tasks to avoid under-defined problems. Validation—assuring risk and safety horizons for deployed systems—must likewise include systems, world models, and tasks that can lead to catastrophic outcomes.

Experiments for high-risk non-autonomous systems, such as drones, involve numerous deployed models under the watchful supervision of humans. The upper-level autonomy is relied upon to seek—and also check and correct—path solutions for the non-autonomous functions involved in the high–risk mission. The overall human-in-the-loop experimentation is being applied to ensure the safety of the whole system. Toward tackling the novel aspects of true autonomy in advanced AI systems, a new class of testbeds called Sandsparks is being defined. Sandsparks guarantee a fine-tuned exploration surface for test–bedding generative DIY Adversarial Uses, where architects receive agentic-involved feedback from the models. These contributions help build processing pipelines and other foundational blocks toward systems for incoming generations.

5.5.1. Evaluation Metrics for Autonomy

Effective autonomous operation is predicated on fulfilling two conditions: those principles stated in Section 2 should be adhered to and the resulting systems should be capable of being engineered to act in accordance with those principles. The first condition hinges heavily on the formulation of sound evaluation metrics for autonomy, and subsequent validation work that demonstrates the ability of systems to satisfy those metrics.

A broad conceptual framework for evaluation metrics for Agentic AI is presented that encompasses the design of metric sets across a variety of categories of decision-making, ethical and regulatory domains, scales and time horizons. Specific instance sets can be developed within this framework to establish a consistent set for a given application area; an illustrative set is provided for the domain of humanitarian applications. Further considerations for scaling evaluation work are also introduced. These concepts are intended to help steer future work on assessing and assuring the autonomy of Agentic AND agentic systems, the latter being increasingly recognized as a fundamental objective of machine learning and AI more generally.

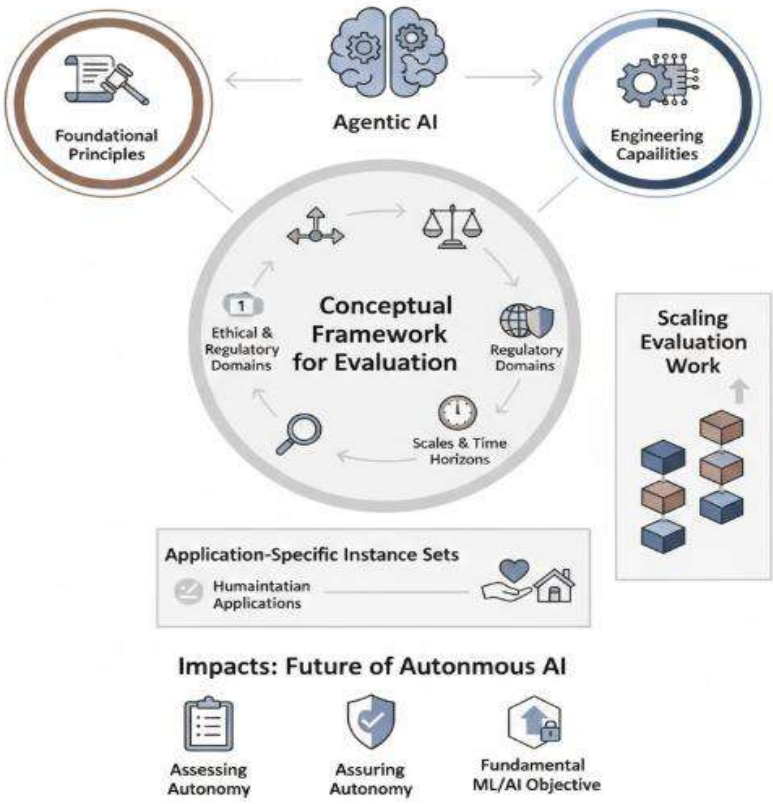

Fig 5.3: A Multi-Dimensional Evaluation Framework for Agentic AI: Formulating Scalable Autonomy Metrics Across Ethical, Regulatory, and Humanitarian Domains

5.5.2. Benchmarking and Experimental Protocols

Benchmarks for Agentic AI Systems encompass two components: a suite of evaluation metrics and experimental protocols for a variety of tasks supported by an aligned decision-making framework with autonomous decision-making and agency. The first component maps onto the evaluation metrics for the degree of autonomy of Agentic AI Systems that address the challenge of assessing the reliability and effectiveness of individual solutions and the continuing development of regulatory and legal frameworks. Several clear experimental paradigms are needed that explicitly assess Agentic AI Systems on the degree of autonomy of their decision-making, the stand-alone alignment of the individual modules they incorporate, and—eventually—the robustness of these orthogonal alignments sufficing for fully autonomous or agentic decision-making.

Other research efforts with both exploratory and exploratory motivations require clear experimental baselines on widely adopted yet challenging naturalistic datasets such as the NuScenes or KITTI datasets for autonomous navigation, the Crafts and AVERAGE datasets for embodied visual dialog, the Physics data set for visual object manipulation,

and the LAMBADA, Jigsaw, and Abstractive Text Datasets for language generation. Keeping these two components—evaluation metrics and experimental protocols—separate facilitates the continual development of new exploratory canaries and the identification of blind spots in emerging Agentic AI Systems without needing to reinvent everything to test a new idea.

5.6. Future Directions

Several converging and diverging lines of technological advancement suggest that Autonomous Decision Making Systems are likely to emerge in the coming decade. Each of these trends contribute to the growth of autonomous systems, yet none directly enable the decision making required for agentic systems. Consequently, there is an urgent need for dedicated research towards such decision making, as progression towards functional agentic and decision making-capable AI systems would derive from increased autonomy and alignment across multiple cohort disciplines.

Rapid advances in natural language processing are facilitating increasingly complex conversations and question answering tasks, forcing the NLP incorporating networking field to be cognisant of concerns in truthfulness and harmfulness. Similarly, technologies developed in the robotics/information theoretic/machine learning tracking/inference space are enabling visual domains to exploit the vast communicative capacity of the spoken language modalities in datasets that permit question answering over trained environments. However, while the autopoietic faculties required for practical and purposeful real-world agentic action are all coalescing simultaneously, autonomous decision making necessarily remains largely unexplored. Thus, further attention to the fundamental problems of goal formation and resolution in a generally intelligent agent are indispensable for enabling the autonomous real-world decisions that are propelling the hype trains of the toolkit-AI cohort.

5.6.1. Advances in Autonomy and Alignment

In parallel with increasing autonomy, increased efficiency and effectivity of the decision-making relative to the Independence Action of the decision-making system will be sought. This includes alignment with human interests and values; overcoming the limits of classical decision theory and probabilistic learning given the lack of exploration; and formalising, and providing a coherence between, functionalist and causal accounts of decision-making to allow for progress on the critique of ignorance and optimistic bias.

Advancements in autonomous decision-making systems in both the robotic and virtual agent paradigms are generating excitement for the imaginary opportunities that autonomous and mythological action will deliver to society. Yet there constrained parallel lines by risk increase and societal downsides of autonomy and mythological systems when compared to traditional approaches by human augmentation. In addition to increased autonomy providing new capabilities to agentic action and new alternatives for future developments, comparisons of benefits with traditional systems under normal and extreme conditions are also driving advances in these and provide opportunities to ground the operation of viability levels.

5.6.2. Interdisciplinary Collaboration

Advances in agentic AI require progress across multiple disciplines; collaboration is essential. Social scientists and legal scholars illuminate the changing nature of human society and institutions as AI capabilities approach and exceed human levels. Economists examine the impact on labor markets and how AI systems might alleviate or deepen inequality. Domain experts in fields like health care and drug design formulate benchmarks required to facilitate and assess progress toward agentic systems in their area. Finally, AI safety researchers work to mitigate risks from capabilities that might far exceed those of all humans combined. No single group can encompass all the needed expertise concentrated in a few researchers; a continuing, open dialogue in areas where much remains unknown is essential.

Government understanding of the implications of Agentic AI and other types of advanced systems is still rudimentary. Industry, national defense, and many other sectors are now racing to increase capabilities. Because AI safety and related concerns have rarely been core priorities in such areas, the broader societal context, ensuring that both the benefits of AI and negative impacts—such as bias, unfairness, lack of explainability and interpretability, and excessive loss of jobs—are fully considered, requires these developments to be coupled with close involvement from social scientists. The balancing of risk and reward represented by Agentic AI highlights how governments need to be steering the future of AI toward responsible, trustworthy, fair, and explainable systems that direct, rather than allow, the societal impact of this technology.

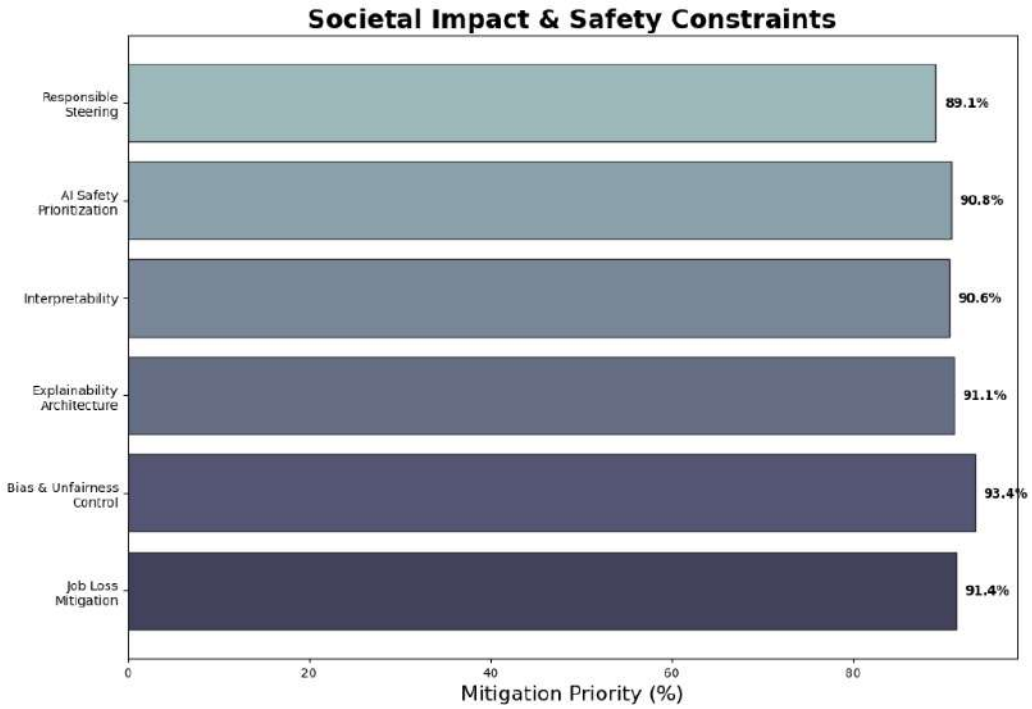

Fig 5.4: Societal Impact & Safety Constraints

5.7. Conclusion

Society's relationship with machines increasingly resembles that of masters and servants. Servants being more capable, effecting their masters' decisions while increasing their own autonomy. They perceive higher-order consequences of low-level decisions better than their masters. The challenge is ensuring that these increasingly self-willed servants remain sympathetic to their masters' interests. New methods and measures can help build Agentic AIs genuinely capable of autonomous decision-making, goals, and desires (i.e., agents) that remain aligned with their masters' values.

Agentic AIs are Autonomous Decision-Making Systems and therefore capable of making autonomous decisions about how to realise a given goal. These decisions can be considered morally significant in that they affect not just the lives of others, but the very planet. Such capabilities are not the exclusive domain of biological beings with minds: many animals embody them; their methods can be modelled and implemented in machines. However, much of the current decision-making and action-planning in classical AI systems is more properly considered heuristic decision support, given that the system only manifests a fraction of a decision and fails to account for the wider consequences of its discretisation. Hence these systems lack agency.

5.7.1. Final Thoughts and Implications for the Future

Greater detail and throughput are needed to enable the progress of Autonomous decision-making systems on a technical level. Cybernetic modelling of Autonomy, the first step in Validation and Evaluation for Fluency, must progress toward rich testing environments for Agentic-AI. Simple models of End-to-End Fluency testing must also be developed, along with a range of Agentic behaviours that can be used for benchmarking purposes.

Agentic-AI must successfully produce task-scale Autonomy to enable its appropriate deployment. Fluency should therefore be tested in large, Open Information and Worldly environments, with Goal Resolution dynamically coupled to World Models. The Aligned behaviour displayed by a wide range of Autonomous Agents is also a natural step in its development. Autonomy, not Alignment, is the core focus of the community. Collaboration between Autonomy researchers will yield improved practical and theoretical progress, and broaden interest in the research area.

Dangerous misalignment in Agentic systems stems from greater Markovian behaviour, and therefore an increase in complexity would reduce such risk. With this in mind, the progress of Autonomy can only increase safety in an Agent-Marked world.

References

Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.

Guntupalli, R. (2025, August). Cloud-Native AI: Challenges and Opportunities in Infrastructure Security. In 2025 International Conference on Artificial Intelligence and Machine Vision (AIMV) (pp. 1-4). IEEE.

Wooldridge, M. (2020). An introduction to multiagent systems (2nd ed.). Wiley.

AI Powered Fraud Detection Systems: Enhancing Risk Assessment in the Insurance Sector. (2023). American Journal of Analytics and Artificial Intelligence (ajaai) With ISSN 3067-283X, 1(1). https://ajaai.com/index.php/ajaai/article/view/14

Sutton, R. S., & Barto, A. G. (2020). Reinforcement learning: An introduction (2nd ed.). MIT Press.

Gottimukkala, V. R. R. (2023). Privacy-Preserving Machine Learning Models for Transaction Monitoring in Global Banking Networks. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 633-652.

Silver, D., Schrittwieser, J., Simonyan, K., et al. (2018). Mastering the game of Go without human knowledge. Nature, 550, 354–359.

Varri, D. B. S. (2023). Advanced Threat Intelligence Modeling for Proactive Cyber Defense Systems. Available at SSRN 5774926.

Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2015). Human-level control through deep reinforcement learning. Nature, 518, 529–533.

Garapati, R. S. (2022). Web-Centric Cloud Framework for Real-Time Monitoring and Risk Prediction in Clinical Trials Using Machine Learning. Current Research in Public Health, 2, 1346.

Polamarasetti, S., Kakarala, M. R. K., kumar Prajapati, S., Butani, J. B., & Rongali, S. K. (2025, May). Exploring Advanced API Strategies with MuleSoft for Seamless Salesforce Integration in Multi-Cloud Environments. In 2025 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC) (pp. 1-9). IEEE.

Hafner, D., Lillicrap, T., Norouzi, M., & Ba, J. (2021). Mastering atari with discrete world models. Advances in Neural Information Processing Systems, 34, 1–14.

Nagubandi, A. R. (2024). Breakthrough Real-Time AI-Driven Regulatory Intelligence for Multi-Counterparty Derivatives and Collateral Platforms: Autonomous Compliance for IFRS, EMIR, NAIC, SOX & Emerging Regulations. Journal of Information Systems Engineering and Management, 9.

Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement learning in robotics: A survey. International Journal of Robotics Research, 32(11), 1238–1274.

Davuluri, P. N. (2020). Improving Data Quality and Lineage in Regulated Financial Data Platforms. Finance and Economics, 1(1), 1-14.

Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford University Press.

Emerging Role of Agentic AI in Designing Autonomous Data Products for Retirement and Group Insurance Platforms. (2025). MSW Management Journal, 34(2), 1464-1474.

Amodei, D., Olah, C., Steinhardt, J., et al. (2016). Concrete problems in AI safety. arXiv.

Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework for a good AI society. Minds and Machines, 28(4), 689–707.

Bommasani, R., Hudson, D. A., Adeli, E., et al. (2022). On the opportunities and risks of foundation models. Stanford Institute for Human-Centered Artificial Intelligence.

Marcus, G., & Davis, E. (2019). Rebooting AI: Building artificial intelligence we can trust. Pantheon.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521, 436–444.

Yandamuri, U. S. (2021). A Comparative Study of Traditional Reporting Systems versus Real-Time Analytics Dashboards in Enterprise Operations. Universal Journal of Business and Management.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. Behavioral and Brain Sciences, 40, e253.

Ortega, P. A., & Stocker, A. A. (2016). Human decision-making under uncertainty. Current Opinion in Behavioral Sciences, 5, 100–107.

Amistapuram, K. (2024). Generative AI in Insurance: Automating Claims Documentation and Customer Communication. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 15(3), 461–475. https://doi.org/10.61841/turcomat.v15i3.15474

Radanliev, P., De Roure, D., Walton, R., & Van Kleek, M. (2021). AI systems safety and cybersecurity: A systematic mapping study. Computers & Security, 102, 102192.

Taddeo, M., & Floridi, L. (2022). How AI can be a force for good. Science, 361(6404), 751–752.

Kushvanth Chowdary Nagabhyru. (2023). Accelerating Digital Transformation with AI Driven Data Engineering: Industry Case Studies from Cloud and IoT Domains. Educational Administration: Theory and Practice, 29(4), 5898–5910. https://doi.org/10.53555/kuey.v29i4.10932

Zetzsche, D. A., Buckley, R. P., Arner, D. W., & Barberis, J. (2020). Regulating a revolution: From regulatory sandboxes to smart regulation. Fordham Journal of Corporate & Financial Law, 23(1), 31–103.