# What can we learn from a billion agents?

## 1 Thesis: Scale, Complexity and Collective Behavior

Imagine a typical morning grocery run in 2030. As you check your phone, your personal AI assistant suggests adding calcium-rich foods to your list, having noted your recent dietary patterns. Your phone gently nudges leaving at 8:45 AM instead of your usual 8:30 - a suggestion being subtly coordinated with hundreds of other shoppers in your neighborhood. As you navigate the store, the layout on your augmented reality display subtly adjusts, not just to optimize your path but to orchestrate a delicate dance of all shoppers, preventing congestion before it occurs. During a surge in local flu cases, your path from door-to-door is automatically adjusted to maintain safe distances from other shoppers, while the store's ventilation systems adapt in real-time to minimize transmission risks.

This isn't science fiction. As billions of AI agents begin to augment human society - managing our calendars, guiding food choices, coordinating healthcare - we're witnessing a massive transformation. This future isn't about replacing human interaction with automation, but enabling unprecedented coordination when we choose to come together. When combined with current technology adoption trends, we are heading to a future where each person could have 10 or more specialized AI agents, adding the equivalent of 60 billion "working agents" to our global systems within a decade. This explosion in artificial agents presents us with a fundamental choice: will these technological capabilities enhance or diminish human autonomy?

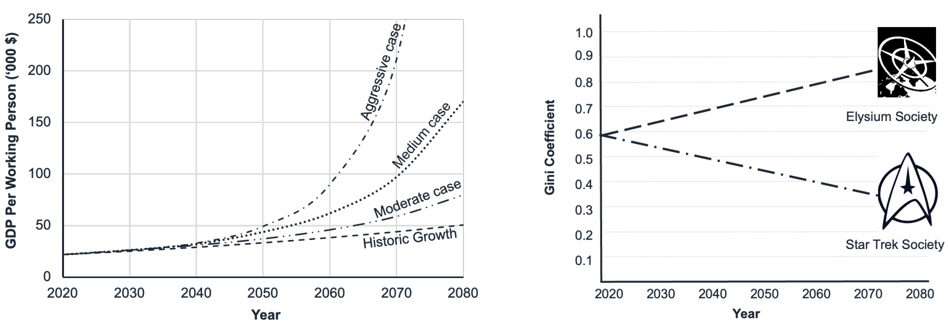


Figure 1: (a) Projected impact of AI agents on global workforce and economic growth, showing acceleration paths to post-scarcity across different adoption scenarios (b) Potential societal outcomes measured by Gini coefficient, contrasting possible futures between an equitable Star Trek-like society and a stratified Elysium-like world. In the age of AI, will we build an Elysium or Star Trek society?

The stakes of this choice manifest vividly in competing visions of our technological future. The Star Trek [20] universe depicts an optimistic 23rd-century where advanced coordination protocols have eliminated material wants, allowing humanity to focus on exploration and advancement. In stark contrast, the film Elysium [1] presents a dystopian 2154 where Earth is overpopulated and impoverished, with wealthy elite enjoying advanced medical technology in an orbital habitat. The tragedy isn't technological limitation - their medical pods could heal any illness - but rather the failure to develop protocols for equitable access and distribution. The difference - Star Trek's abundance

Preprint. Under review.

versus Elysium's scarcity - lies not in technological capability, but in their ability to coordinate human activity and resources at unprecedented scales.

Yet, throughout history, humans have been limited to meaningfully maintaining a few hundred stable relationships - a constraint known as Dunbar's number [12, 13]. Even in our hyper-connected digital age, while we can theoretically access millions through social media, our cognitive architecture remains fundamentally limited. As our technological reach expands exponentially through billions of AI agents, our human capacity to understand and guide these interactions remains fixed. The solution, however, lies not in making individual agents smarter or in replicating human cognition, but in discovering protocols that enable beneficial coordination at scales far beyond human cognitive limits - protocols that enhance rather than replace human agency.

From our morning grocery run to global pandemic response, the future of human flourishing depends on our ability to coordinate effectively at unprecedented scales while preserving individual autonomy. This requires a fundamental shift - from model-centric systems aimed at enhancing cognitive capabilities to protocol-centric systems that discover unprecedented coordination.

## 2 Nature Protocols and Human Coordination

Nature offers profound insights into solving massive-scale coordination challenges [14]. Consider how army ants construct living bridges: each ant follows a remarkably simple protocol - if there's an ant in front, cross the bridge; if not, become part of the bridge. This minimal set of rules, when executed by thousands of ants simultaneously, creates remarkably resilient structures that no individual ant could comprehend or design. A back-of-the-envelope calculation reveals that the strength-to-weight ratio of an ant bridge is 1,000-24,000 times higher than that of a human-built concrete bridge [22]. This simple example illustrates the fundamental power of protocols - rules of interaction that transform local actions into sophisticated collective behavior.

Natural systems demonstrate increasingly complex coordination challenges. In multicellular organisms, evolution has discovered reliable protocols for coordinating millions of cells through shared genetic instructions. Fish schools achieve sophisticated collective movement through simple interaction rules that balance individual survival with group benefits. Bacterial colonies demonstrate how even organisms with limited individual capabilities can achieve remarkable collective outcomes through carefully evolved protocols. These systems show nature's solution to a fundamental challenge: how to achieve sophisticated collective behavior through simple, local interaction rules even when individual and group interests partially conflict [19, 24, 11].

Our greatest societal challenges mirror this need for coordination beyond individual capability. The COVID-19 pandemic showed how individual outcomes depend not on any single decision but on millions of interlinked choices about testing, isolation, and vaccination. Here, individual incentives for normal social interaction conflicted with collective needs for isolation and distancing. Climate change presents similar dynamics, where individual benefits from carbon emissions create collective harm. These aren't failures of individual decision-making, but our inability to coordinate at scale.

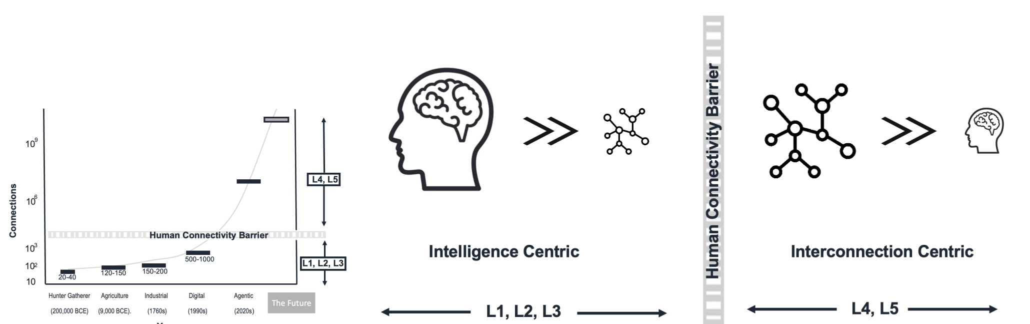


Figure 2: (a) The Human Connectivity Barrier: cognitive limits of human recognition and relationship maintenance (b) Evolution of agentic systems from operating within human cognitive bounds to enabling coordination beyond these natural limits through protocol-centric approaches

Throughout history, humans have developed increasingly sophisticated protocols to transcend these limitations. Early writing systems enabled asynchronous coordination across time and space. Railway timetables synchronized society across vast distances. The Internet's TCP/IP protocols [4, 10] enabled global information exchange without central control. Yet each advance, while expanding our collective reach, hit the same fundamental barrier: human cognitive limits. Medieval guild masters could train only a handful of apprentices. Railway operators could manage only a limited number of routes. Even in today's digital age, while social media theoretically connects billions, meaningful engagement remains bounded by Dunbar's number - our cognitive limit of around 150 stable relationships and 1,500 recognizable faces [1].

This constraint isn't just about individual relationships; it fundamentally limits the complexity of protocols themselves. Traditional protocols must remain comprehensible to the humans who design and oversee them. Even sophisticated systems like financial markets or supply chains must ultimately operate within human cognitive bounds. This creates a crucial tension: as our technological reach expands exponentially through billions of AI agents, our human capacity to understand and guide these interactions remains fixed.

The imminent emergence of billions of AI agents presents an unprecedented opportunity to transcend these cognitive limits entirely. Unlike historical protocols that had to remain "human-readable", agentic protocols can operate at complexities and scales far beyond human comprehension while still producing beneficial outcomes. When agents execute protocols on our behalf, they can simultaneously process millions of information sources, maintain complex relationships with thousands of other agents, and coordinate actions across massive networks while adapting to real-time feedback. This represents more than a quantitative expansion of existing capabilities - it offers a qualitative shift in how we approach collective coordination.

This progression reveals why traditional AI approaches, focused on enhancing individual agent intelligence, hit limits when addressing collective challenges. Even sophisticated language models excel at processing information and making individual decisions, but they cannot inherently solve problems requiring massive-scale coordination. The path forward requires a shift in focus - from making smarter agents to enabling smarter interactions. The future of AI is protocol-centric.

## 3 Path to Protocol-Centric Intelligence

Current artificial intelligence, dominated by Large Language Models [3] and multi-agent systems, has dramatically enhanced individual capabilities - from writing assistance to autonomous task completion. However, these systems fundamentally operate within human cognitive bounds, limited to direct person-to-person interactions or small networks of specialized agents [25, 23]. While an LLM-powered agent can expertly manage your calendar or a multi-agent system can coordinate a team meeting, they cannot address challenges that emerge from millions of simultaneous interactions [5].

The solution lies not in making individual agents smarter, but in discovering protocols that enable beneficial coordination at scales far beyond human cognitive limits. Just as ant colonies achieve remarkable feats through simple interaction rules rather than individual ant intelligence, we need systems that can discover and implement sophisticated coordination protocols at societal scales. Large Population Models (LPMs) provide the mathematical and computational framework for this protocol-centric approach. They rest on three fundamental pillars:

- **Synthetic Agent Networks**: LPMs creating "computational petri dishes" where millions of agents interact and evolve together [8]. Unlike simplified simulations, these synthetic environments maintain rich agent behaviors - from basic heuristics to sophisticated language model-powered interactions [5]. This enables modeling complex phenomena that emerge only at population scale: how individual commuting decisions create traffic patterns, how local interactions shape disease transmission, or how small disruptions cascade through supply chains. Just as biological evolution discovered effective protocols through repeated interaction and selection, these synthetic environments enable systematic discovery of coordination strategies that would be impossible to design manually.

[1] https://www.theatlantic.com/family/archive/2021/05/robin-dunbar-explains-circles-friendship-dunbars-number/618931/

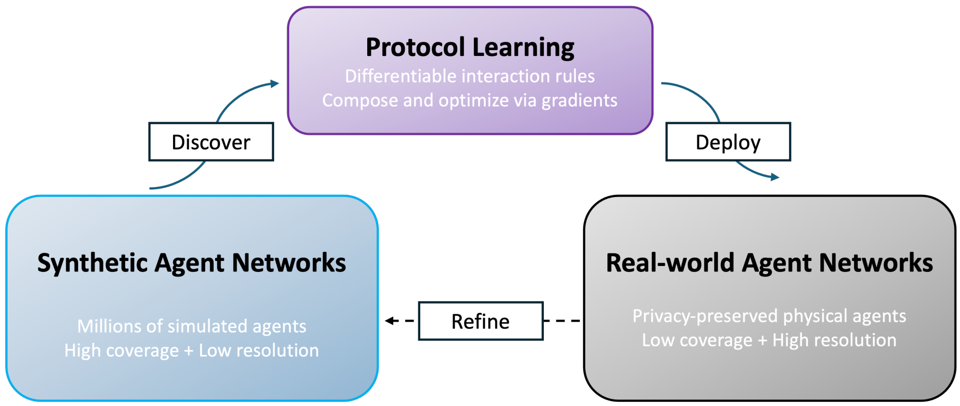


Figure 3: **Large Population Models enable bidirectional learning between synthetic and physical agent networks through differentiable protocols**. Synthetic networks simulate millions of agents (e.g., modeling disease spread in a city) with high population coverage but lower granularity. These simulations discover initial interaction protocols through gradient-based optimization. The learned protocols are then deployed to real-world agent networks (e.g., privacy-preserved contact tracing systems) that provide high-resolution ground truth but with limited coverage. Because the entire system remains differentiable, protocols can be continuously refined using real-world feedback while maintaining privacy - effectively allowing "backpropagation through reality." This creates a learning loop where protocols improve through broad synthetic exploration and precise real-world validation.

- **Differentiable Protocols**: LPMs make these massive agent behaviors and interactions differentiable - enabling systematic discovery and refinement of coordination protocols through gradient-based optimization [7]. When a simulation reveals suboptimal collective outcomes, the system can automatically trace how small changes in interaction rules affect population-level results. This isn't just about faster simulation - it enables automatic discovery of coordination protocols that might be unintuitive to human designers but provably effective at scale. Most importantly, maintaining end-to-end differentiability creates a crucial bridge between synthetic and physical agents, enabling continuous learning from both simulated and actual behavioral data.
- **Physical Agent Networks**: The ultimate test of protocol effectiveness lies in real-world deployment. Through novel cryptographic approaches, LPMs enable actual people, devices, and organizations to participate directly in protocol refinement while preserving privacy [6]. This creates a living feedback loop: protocols discovered through synthetic simulation deploy to physical agents, whose behaviors then inform further protocol evolution. When your phone suggests leaving for groceries five minutes later than usual, it's participating in a distributed protocol network that's continuously learning and adapting to optimize collective outcomes while preserving individual privacy and choice.

This approach represents a fundamental advance over existing coordination approaches. Traditional mechanism design focuses on creating static rules for small groups [2, 18], while modern multi-agent protocols like MCP and AgentProtocol provide standardized APIs for direct task coordination. Yet many of our pressing societal challenges require coordinating millions of individuals through protocols that can continuously adapt to changing conditions. LPMs transcend these limitations in three key dimensions refine bullets:

- **Scale**: While democratic AI [15] approaches enable learning adaptive mechanisms through reinforcement learning, they remain constrained to small, synthetic environments (2-10 agents). Recent work shows that protocols coordinating millions of agents requires capturing population-scale, as indirect network effects dominate behavior [5]. LPMs enable discovery and validation of protocols that can handle these population-scale dynamics where individual behaviors and incentives shift based on the actions of millions of others.
- **Data**: Modern multi-agent systems [25, 23] are constrained to small, closed groups (10-1000 agents) with direct interactions networks. Beyond scale challenges, they rely on unconstrained natural language protocols that cannot be systematically calibrated against

real-world data. LPMs enable protocols to be flexibly specified (neural, mechanistic, generative etc) validated through large-scale simulations and continuously refined based on observed behaviors and outcome data.

- **Sim2Real Transfer**: While previous work on 'society-in-the-loop' design [21] and decentralized mechanism design [17] highlighted the importance of real-world feedback for protocol refinement, they lack concrete paths to operationalization. LPMs bridge this gap through parallel simulation of synthetic and physical agent networks, enabling continuous protocol refinement while preserving privacy through decentralized computation.

Unlike current frameworks that rely on explicit communication between agents, this protocol-centric approach enables beneficial collective behavior to emerge naturally from local interactions, similar to how ant colonies achieve sophisticated coordination without centralized control. The framework transforms how we approach societal challenges. Rather than building increasingly sophisticated "digital humans", we create the computational infrastructure for "digital nations" - where millions of agents, both synthetic and real, can coordinate effectively while preserving individual privacy and agency. This enables progression through increasingly sophisticated coordination capabilities (detailed in the next section) - from personal assistance to global orchestration - while maintaining individual autonomy. Just as TCP/IP protocols enabled global information exchange without central control, LPM-discovered protocols may enable unprecedented human coordination while preserving individual choice and privacy.

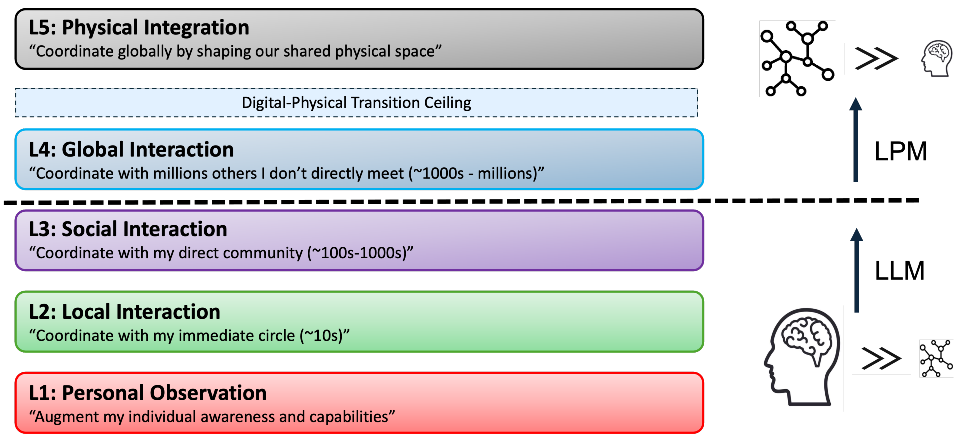


Figure 4: The Individual-Interconnection Transition: From Local to Global Context in Individual Decision-Making. Lower levels (L1-L3) are LLM-based and focus on enhancing individual and small-group capabilities. Higher levels (L4-L5) are LPM-based to enable coordination beyond human cognitive limits by orchestrating millions of simultaneous interactions - in digital (L4) and physical worlds (L5).

## 4 The Five Levels of Agentic Systems

AI agents will evolve to dramatically expand human agency - enabling each individual to understand and shape increasingly wider circles of influence. We propose a five-level framework that traces this expansion of personal impact, from enhancing individual capabilities to shaping physical reality. This framework illuminates both the immediate benefits to individuals and the underlying technological advances needed at each stage.

To ground our discussion, we follow two individuals navigating complex challenges through increasingly sophisticated AI agents: Sarah Chen, a restaurant owner in Kansas City working to keep her business and community safe during a pandemic, and Michael Roberts, a software engineer seeking to optimize not just his own but thousands of others' festival experience at Coachella.

**Level 1 (Personal Lens - Enhancing Individual Awareness)** agents enhance your individual understanding through information synthesis. Using retrieval-augmented generation [16], these agents process and summarize complex information without requiring any coordination. Sarah's

COVID analyst agent distills epidemiological data into clear insights, while Michael's festival expert summarizes years of Coachella experiences. This represents the current state of personal AI assistance, where agents excel at information processing but operate independently.

**Level 2 (Local Interaction - Managing Immediate Circle)** enables coordination between your own AI agents through standardized protocols. Emerging standards like Model Context Protocol [2] allows your agents to seamlessly share information and coordinate actions. Sarah's agents coordinate internally - her inventory agent communicates with her scheduling agent to automatically order cleaning supplies based on staff rotations. Michael's travel agent coordinates with his calendar and budget agents to book optimal flights and accommodations. This internal orchestration optimizes your personal workflow through reliable agent-to-agent communication.

**Level 3 (Social Coordination - Coordinating Bounded Networks)** extends coordination to bounded networks of other people's agents. Frameworks like AutoGen [25] and AgentProtocol[3] enable small-group orchestration with clear protocols. Sarah's staff scheduling agent coordinates directly with her employees' calendar agents to optimize shift coverage. Michael's arrival agent synchronizes with his friends' agents to coordinate festival meetup times. Current multi-agent systems excel at this direct, small-group coordination, though they remain limited to explicit, observable interactions.

**Level 4 (Global Orchestration - Understanding Population Influence)** marks a fundamental shift - similar to how autonomous vehicles must consider cascading traffic patterns rather than just immediate surroundings. When Sarah adjusts her restaurant's testing schedule, it affects not just her immediate staff but creates ripple effects through customer and community networks. These indirect effects are computationally irreducible, requiring LPMs' synthetic agent networks to discover optimal coordination protocols through massive-scale simulation using AgentTorch [9]. Similarly, Michael's movement choices at Coachella influence crowd flow through complex network effects that only emerge at population scale.

**Level 5 (Physical Integration - Shaping Collective Realtiy)** moves from understanding these network effects to actively shaping them in real-time. Your choices help evolve protocols that orchestrate physical spaces while preserving individual autonomy. Michael's movement preferences contribute to protocols that naturally prevent congestion through decentralized coordination, while Sarah's operational decisions help shape dynamic "immunity networks" where testing stations and ventilation systems adapt to emergent patterns. This requires LPMs to bridge synthetic protocol discovery with privacy-preserved physical agent networks.

The key transition between L3 and L4 marks a shift from direct to indirect influence. While earlier levels coordinate through explicit agent-to-agent communication, L4 and L5 capture complex network effects that emerge from millions of interconnected decisions. This requires the technical foundations of LPMs: synthetic networks discover potential protocols, differentiable protocols enable systematic refinement, and physical networks allow privacy-preserved real-world deployment.

This progression represents a fundamental advance in human coordination. Rather than simply extending direct influence, the higher levels enable you to participate in discovering entirely new ways for populations to coordinate effectively. The transition from understanding network effects through simulation (L4) to shaping them in physical reality (L5) moves us from reactive to adaptive coordination, all while preserving individual agency. The future of AI lies not in replacing human decision-making, but in expanding each person's ability to positively influence our collective behavior.

## 5 L4-L5 Transition: From Understanding to Shaping Behavior

While the progression through Levels 1-4 represents expanding circles of influence, each level fundamentally operates in the digital realm - discovering protocols to coordinate among AI agents. Powered by LPMs' ability to enable agents to coordinate millions of digital interactions, Level 5 marks a transformative leap: extending this orchestration into physical reality. The fundamental distinction between Level 4 and Level 5 lies not in the number of coordinated interactions, but in their nature - from coordinating digital information flows to actively shaping physical protocols themselves. This isn't just about processing more data or expanding coordination scope - it's about transforming

[2]https://modelcontextprotocol.io/introduction
[3]https://agentprotocol.ai/

how artificial intelligence enables individual agents to harmoniously interact with and shape physical reality.

This distinction becomes clear through our examples. At Level 4, Sarah's restaurant management agent can process real-time pandemic data and interactions of millions to recommend optimal testing strategies. But it's limited to prediction and recommendation - when it suggests using rapid tests despite lower accuracy, it's because that's the best coordination possible given how others might behave. At Level 5, the agent doesn't just predict behavior - it actively shapes physical interactions. Testing stations dynamically adjust their operations, ventilation systems create coordinated airflow patterns, and customer devices form living "immunity networks" that prevent transmission hotspots before they form.

This evolution mirrors the evolution of navigation technology with self-driving vehicles. Traditional navigation apps (with current Level 4 vehicles) can analyze traffic data from millions of vehicles to suggest optimal routes to individual drivers. But being limited to digital coordination, they often create "ghost traffic jams" when too many vehicles follow similar recommendations. Yet at Level 5 (as full self-driving emerges), each driver's choices will become part of a living protocol network - their individual routing decisions subtly influencing traffic signals, lane configurations, and other drivers' suggestions in real-time, creating naturally flowing traffic patterns without central control.

| Dimension | State |
|---|---|
| Level 1: Personal | Stable |
| Level 2: Local | Current |
| Level 3: Social | Current |
| Level 4: Global | Emerging |
| Level 5: Physical | Future |

(a) Technology Readiness for agentic systems

| Dimension | Level 4: Orchestrate | Level 5: Harmonize |
|---|---|---|
| Primary Domain | Digital environments | Physical reality |
| Individual's Role | System observer | System shaper |
| Decision Support | Shows possible outcomes | Guides collective evolution |
| Infrastructure | Works within existing systems | Creates new protocols |
| Real-world Impact | Through recommendations | Through direct integration |
| Feedback Loop | One-way learning | Continuous adaptation |

(b) The capability transition from L4 to L5 systems

Figure 5: Technology readiness and capability transition from L4 to L5 systems: from predictive recommendations to active physical-world orchestration.

Similarly, Michael's festival experience demonstrates this transition from digital to physical orchestration. At L4, his app processes crowd movement data to suggest optimal viewing locations. At L5, his individual choices about when and where to move become part of a dynamic physical choreography - his device participating in subtle peer-to-peer protocols that naturally prevent congestion while preserving his freedom to explore the festival as he wishes.

The key breakthrough enabling this transition is the protocol-centric design of LPMs. Unlike traditional AI that focuses on enhancing individual decision-making in isolation, LPMs discover protocols that allow individual choices to naturally harmonize through local interactions. They maintain privacy and agency while enabling unprecedented coordination across both digital and physical domains.

# 6 Risks and Net Benefits of Protocol-Centric Intelligence

The development of protocol-centric intelligence presents three fundamental challenges that must be carefully addressed:

- **Privacy and Control**: While these systems enable decentralized coordination, they could potentially enable unprecedented surveillance. A system designed to coordinate shopping patterns could track movements; one designed for pandemic response could enable unwanted social monitoring. The solution lies in privacy-preserving computation and decentralized protocols that make such abuse technically impossible, not just prohibited.
- **Digital Divide 2.0**: Without careful design, these coordination capabilities could create "protocol privilege" - where those with access to advanced protocols optimize their lives at the expense of others. Similar to how internet access defines opportunities today, protocol access could become a new axis of inequality.

- **The Agency Paradox**: In pursuing collective efficiency, poorly designed systems could subtly restrict individual choice. The challenge is maintaining true optionality while enabling coordination benefits.

However, the potential benefits of protocol-centric intelligence are transformative:

- **Public Health**: Rather than choosing between individual freedom and collective safety, adaptive protocols could enable fine-grained pandemic responses that maintain both. Privacy-preserving contact tracing networks could provide early warnings while protecting individual privacy.
- **Climate Action**: Instead of relying on regulations or individual sacrifice, these systems could discover protocols that make sustainable choices naturally advantageous. From optimizing shared transportation to coordinating energy use, they could help address climate change while improving quality of life.
- **Urban Living**: Cities could become living systems that adapt to inhabitants' needs, preventing congestion before it forms and dynamically adjusting public spaces to enhance density without sacrificing livability.

The net positive impact emerges from three key principles: i) Technical decentralization that prevents central control, ii) Enhancement of human agency rather than replacement, iii) Universal access by design to reduce inequalities.

## 7 Conclusion: Choosing Our Protocol-Centric Future

Return to our opening contrast: Star Trek's abundance through coordination versus Elysium's scarcity through fragmentation. The difference lies not in raw technological capability - both societies had advanced AI - but in their ability to coordinate human activity at scale while preserving individual agency. As billions of AI agents enter our society, we face the same choice. Will these agents enhance our autonomy by enabling unprecedented coordination, or diminish it through fragmentation and control? The answer lies not in making individual agents smarter, but in discovering protocols that enable beneficial collective behavior while preserving individual choice.

Our morning grocery run in 2030 could reflect either future. In one, AI agents subtly coordinate millions of individual choices to prevent congestion, reduce waste, and enhance everyone's experience. In another, these same technologies create invisible barriers, privileging some while excluding others. The difference lies in the protocols we discover and implement today.

The future of artificial intelligence depends not on replicating human cognition, but on enabling unprecedented human coordination. By focusing on protocol-centric intelligence - built on principles of decentralization, enhancement, and universal access - we can ensure that technological advancement serves to expand rather than constrain human potential. The choice between abundance through coordination and scarcity through fragmentation isn't predetermined - it's a protocol we get to write.

## References

[1] Elysium. Wikipedia, 2013.

[2] Rediet Abebe and Kira Goldner. Mechanism design for social good. *AI Matters*, 4(3):27–34, 2018.

[3] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.

[4] Vinton G Cerf and Robert E Kahn. A protocol for packet network intercommunication. *IEEE Transactions on Communications*, 22(5):637–648, 1974.

[5] Ayush Chopra, Shashank Kumar, Nurullah Giray-Kuru, Ramesh Raskar, and Arnau Quera-Bofarull. On the limits of agency in agent-based models. In *Proceedings of the 24th International Conference on Autonomous Agents and Multi-Agent Systems*, AAMAS '25, 2025.

[6] Ayush Chopra, Arnau Quera-Bofarull, Nurullah Giray-Kuru, Michael Wooldridge, and Ramesh Raskar. Private agent-based modeling. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multi-Agent Systems*, AAMAS '24, 2024.

[7] Ayush Chopra, Alexander Rodríguez, Jayakumar Subramanian, Arnau Quera-Bofarull, Balaji Krishnamurthy, B Aditya Prakash, and Ramesh Raskar. Differentiable agent-based epidemiology. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multi-Agent Systems*, AAMAS '23, 2023.

[8] Ayush Chopra, Jayakumar Subramanian, Balaji Krishnamurthy, and Ramesh Raskar. flame: A framework for learning in agent-based models. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, pages 391–399, 2024.

[9] Ayush Chopra and team. AgentTorch: Large population models, 2024.

[10] David D Clark. The design philosophy of the darpa internet protocols. *ACM SIGCOMM Computer Communication Review*, 18(4):106–114, 1988.

[11] Iain D Couzin, Jens Krause, Nigel R Franks, and Simon A Levin. Effective leadership and decision-making in animal groups on the move. *Nature*, 433(7025):513–516, 2005.

[12] Robin I M Dunbar. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6):469–493, 1992.

[13] Russell A Hill and Robin I M Dunbar. Social network size in humans. *Human Nature*, 14(1):53–72, 2003.

[14] Bert Hölldobler and Edward O Wilson. *The Ants*. Harvard University Press, 1990.

[15] Raphael Koster, Jan Balaguer, Andrea Tacchetti, Ari Weinstein, Tina Zhu, Oliver Hauser, Duncan Williams, Lucy Campbell-Gillingham, Phoebe Thacker, Matthew Botvinick, et al. Human-centred mechanism design with democratic ai. *Nature Human Behaviour*, 6(10):1398–1407, 2022.

[16] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.

[17] Jin Xing Lim, Barnabé Monnot, and Georgios Piliouras. Blockchain-based mechanism design for collaborative mathematical research. In *2022 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, pages 1–9. IEEE, 2022.

[18] Eric S Maskin. Mechanism design: How to implement social goals. *American Economic Review*, 98(3):567–576, 2008.

[19] Richard E Michod. Evolution of individuality during the transition from unicellular to multicellular life. *Proceedings of the National Academy of Sciences*, 104(suppl 1):8613–8618, 2007.

[20] Michael Okuda and Denise Okuda. *The Star Trek Encyclopedia*. Simon and Schuster, 2011.

[21] Iyad Rahwan. Society-in-the-loop: programming the algorithmic social contract. *Ethics and information technology*, 20(1):5–14, 2018.

[22] Chris R Reid, Matthew J Lutz, Scott Powell, Albert B Kao, Iain D Couzin, and Simon Garnier. Army ants dynamically adjust living bridges in response to a cost-benefit trade-off. *Proceedings of the National Academy of Science*, November 2015.

[23] Alexander Sasha Vezhnevets, John P Agapiou, Avia Aharon, Ron Ziv, Jayd Matyas, Edgar A Duéñez-Guzmán, William A Cunningham, Simon Osindero, Danny Karmon, and Joel Z Leibo. Generative agent-based modeling with actions grounded in physical, social, or digital space using concordia. *arXiv preprint arXiv:2312.03664*, 2023.

[24] Stuart A West, Ashleigh S Griffin, Andy Gardner, and Stephen P Diggle. Social evolution theory for microorganisms. *Nature Reviews Microbiology*, 4(8):597–607, 2006.

[25] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Shaokun Zhang, Erkang Zhu, Beibin Li, Li Jiang, Xiaoyun Zhang, and Chi Wang. Autogen: Enabling next-gen llm applications via multi-agent conversation framework. *arXiv preprint arXiv:2308.08155*, 2023.