EPINOVA–2026–PB–59

Policy Brief

Series Information:
This policy brief is part of the EPINOVA Policy Brief Series on Strategic Competition, AI-Enabled Warfare, and Information Conflict.

Recommended Citation:
Wu, Shaoyuan (2026), *AI Capability Stratification: A Framework for the Future Distribution of AI Power*, Policy Brief No. EPINOVA–2026–PB–59, Global AI Governance and Policy Research Center, EPINOVA LLC.
https://doi.org/10.67037/epinova.pb.2026.059

Disclaimer:
This policy brief is an institutional publication of EPINOVA, prepared by Dr. Shaoyuan Wu in his capacity as Director of the Global AI Governance and Policy Research Center, EPINOVA LLC. The analysis is based on publicly available information and scenario-based analytical estimates and does not represent the official position of any government. The publication is intended solely for research and policy discussion purposes and does not constitute legal, military, operational, or sanctions-compliance advice.

GLOBAL AI GOVERNANCE
RESEARCH CENTER

EPINOVA

AI Capability Stratification:

A Framework for the Future Distribution of AI Power

Author: Shaoyuan Wu
Affiliation: Global AI Governance and Policy Research Center, EPINOVA LLC
Date: June 26, 2026

Key Judgments

- **Public AI is only the visible surface of future AI power.** Public-facing assistants, copilots, coding agents, and generative tools shape mass adoption, but they do not represent the full structure of AI power.
- **AI capability will stratify across access, data, authority, specialization, interface, and rule space.** Future AI systems will differ not only by model performance, but also by who can use them, what they can access, what they can influence, and under what rules they operate.
- **Deep AI will become a primary layer of professional productivity.** The most durable value is likely to come from AI embedded in organizations, proprietary knowledge systems, expert workflows, and operational decision processes.
- **Dark-domain AI should be understood structurally, not morally.** It refers to restricted visibility, high access thresholds, concentrated authority, resource connection, real-world consequence, and differentiated rule conditions.
- **AGI would not eliminate stratification.** Even highly general AI would not automatically possess all proprietary data, institutional memory, expert feedback, resource authority, operational interfaces, or special permissions.
- **AI governance must move beyond public-model regulation.** Governance will need to address embedded, operational, proprietary, high-authority, and boundary-condition AI systems.

Executive Summary

The future development of artificial intelligence should not be understood as a linear race among public-facing large models. As AI systems become embedded in institutions, workflows, operational systems, proprietary knowledge environments, resource-allocation mechanisms, and differentiated rule conditions, AI capability is likely to form a stratified structure.

This policy brief proposes **AI Capability Stratification** as a structural framework for understanding the future distribution of artificial intelligence power. The framework shifts attention from model performance alone to the broader conditions under which AI capability is accessed, embedded, authorized, optimized, and connected to real-world systems.

The model divides AI capability into three broad domains: **Surface Intelligence Domain**, **Deep Intelligence Domain**, and **Dark-Domain Intelligence**. These domains can be further divided into seven analytical layers: **Public Interface Intelligence**, **Institutional Embedded Intelligence**, **Operational Control Intelligence**, **Proprietary Epistemic Intelligence**, **Objective-Specific Optimized Intelligence**, **High-Authority Resource Intelligence**, and **Boundary-Condition Intelligence**.

The central thesis is straightforward: **Public AI determines visible access; Deep AI determines professional productivity; and Dark-domain AI determines how AI capability is linked to concentrated authority, resource command, real-world consequence, and differentiated rule conditions.**

This framework does not claim that deeper AI layers are always more intelligent, more advanced, or more dangerous. Rather, it argues that AI power depends on the interaction between model capability, data quality, access conditions, institutional embedding, resource authority, real-world interfaces, and rule space.

The future distribution of AI power will therefore not be determined only by which model is the strongest. It will also be determined by who can use AI, what AI can access, what resources AI can influence, what systems AI can affect, and under what rules AI is allowed to operate.

Why This Matters

AI policy debates still focus heavily on public-facing models, benchmark performance, and consumer applications. These issues matter, but they capture only the most visible layer of AI capability. As AI systems become embedded in institutions, operational workflows, proprietary knowledge environments, and high-authority decision structures, the distribution of AI power will increasingly depend on access, data, authority, resources, interfaces, and rule space.

The most consequential AI systems may not be the most visible ones. A public assistant may shape mass adoption, but an embedded system connected to proprietary data, operational decisions, or resource-allocation authority may have greater institutional, economic, or strategic impact.

For policymakers, the implication is clear: AI governance must move beyond public-model regulation. Effective oversight will require attention to where AI systems are embedded, what they can access, what decisions they can influence, what resources they can command, and under what rules they operate.

Analytical Note

This policy brief uses AI power to refer to the system-level capacity of AI to shape decisions, allocate resources, affect institutions, or intervene in real-world processes once connected to data, authority, infrastructure, operational interfaces, and rule conditions. AI power is therefore a system-level property, not merely a model-level property (Wu, 2026).

The framework is structural rather than predictive. It does not claim that all AI systems will move through the seven layers in sequence, or that deeper layers are necessarily more intelligent, more advanced, or more dangerous. The layers describe different configurations of access, authority, specialization, resource connection, and rule space.

Policy Brief

The terms **Surface Intelligence**, **Deep Intelligence**, and **Dark-Domain Intelligence** are used as analytical categories. They should not be read as moral labels. In particular, **Dark-Domain Intelligence** does not imply illegality, malicious intent, or inherent opacity. It refers to AI systems operating under restricted visibility, high access thresholds, concentrated authority, significant resource connection, real-world consequence, and differentiated rule conditions.

1. The AI–Internet Stratification Analogy

The future distribution of artificial intelligence power can be understood through a structural analogy with the internet.

The public web is visible, searchable, and broadly accessible. The deep web consists of permissioned, database-driven, account-based, institutional, or internal systems that are not publicly indexed. The dark web refers to low-visibility, high-access-threshold environments with distinct identity, access, and rule structures (Bergman, 2001; Gupta et al., 2021).

Applied to AI, this analogy is structural rather than moral. It does not distinguish good AI from bad AI, legal AI from illegal AI, or transparent AI from black-box AI. Instead, it highlights differences in visibility, access, authority, specialization, operational connection, and rule conditions.

In AI, the relevant distinctions are between mass-audience systems and restricted-audience systems, open-access tools and permissioned tools, general-purpose models and objective-specific systems, low-resource-authority applications and high-resource-authority systems, public-rule environments and differentiated-rule environments, and cognitive-assistance tools and systems capable of real-world intervention.

Just as the public web does not represent the full internet, public-facing AI does not represent the full structure of future AI capability. Public AI may dominate visibility, public debate, consumer adoption, and media attention. However, the deeper distribution of AI power will depend on less visible systems embedded in institutions, markets, infrastructures, security environments, and strategic decision-making structures.

The analogy also clarifies a common analytical error: visibility should not be confused with importance. A public AI assistant may be more widely used, but an embedded AI system connected to proprietary data, operational workflows, resource-allocation authority, or real-world execution interfaces may be more consequential.

The purpose of the analogy is therefore not to reproduce the internet's categories inside AI. It is to show that public visibility captures only one layer of a larger capability structure. Future AI power will be distributed across multiple domains of access, knowledge, authority, resources, interfaces, and rule space.

2. Structural Variables of AI Capability Stratification

AI Capability Stratification is defined by seven structural variables: **audience scale, access threshold, data quality, domain specificity, resource authority, real-world intervention capacity,** and **rule space**. Together, these variables explain how AI capability is distributed, constrained, amplified, and translated into power.

Audience scale concerns who is authorized to use an AI system. Public-facing systems serve mass audiences, while restricted systems may serve professional communities, firms, agencies, military organizations, research institutions, or small authorized groups. **Access threshold** concerns the technical, institutional, financial, legal, or security barriers required to use a system. Higher access thresholds reduce public visibility but may increase specialization, control, and consequence.

Data quality refers to the relevance, exclusivity, reliability, timeliness, and domain value of the data available to an AI system. In many cases, data quality and exclusivity may matter more than model scale. **Domain specificity** refers to the extent to which a system is optimized for a particular sector, function, environment, task, or decision problem.

Resource authority refers to the extent to which an AI system can recommend, allocate, trigger, or command material, financial, computational, organizational, legal, administrative, or coercive resources. **Real-world intervention capacity** refers to whether the system remains advisory or can directly affect workflows, machines, markets, infrastructure, administrative systems, military systems, or other external environments. These distinctions align with broader AI governance concerns about system context, intended use, impact pathways, and risk management (National Institute of Standards and Technology [NIST], 2023; OECD, 2019; Shevlane et al., 2023).

Rule space refers to the legal, regulatory, organizational, ethical, and exceptional conditions under which an AI system operates. Public consumer AI operates under one rule space, while enterprise, defense, intelligence, emergency, research, and critical-infrastructure systems may operate under others.

These variables are mutually reinforcing. Smaller audiences often correspond to higher access thresholds. Greater domain specificity usually depends on higher-quality data. Stronger resource authority often increases real-world intervention capacity. Differentiated rule conditions may also enable forms of capability unavailable to public AI.

The result is a layered distribution of AI power. AI capability is not merely a property of the model. It is a property of the model embedded within a specific data environment, authority structure, resource system, interface layer, and rule regime.

3. Three Domains of AI Capability

Figure 1 summarizes the full framework before the brief turns to the three domains and seven layers in detail.

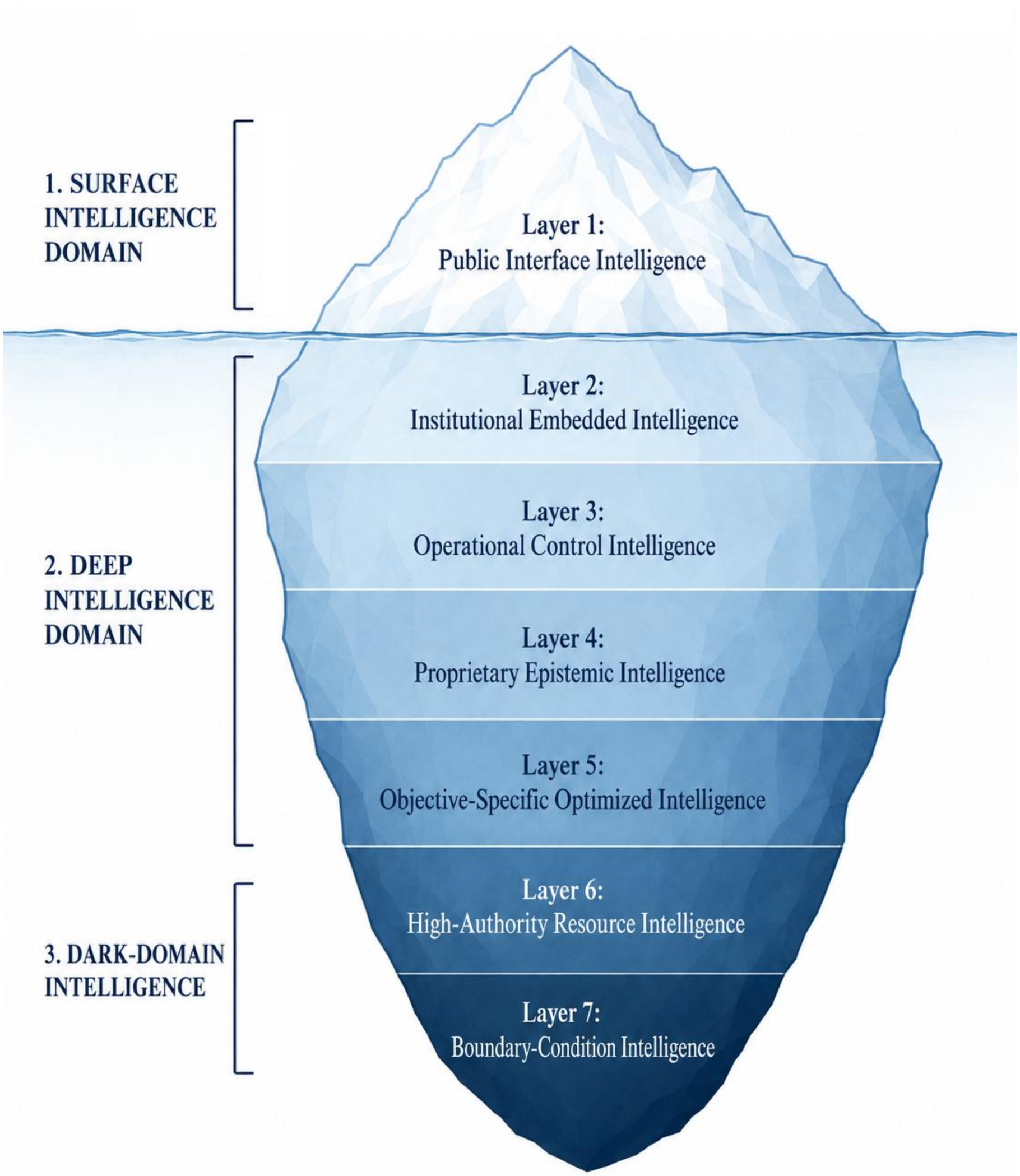

Figure 1. AI Capability Stratification: Three Domains and Seven Layers

The figure presents the AI Capability Stratification framework as an iceberg model. The visible surface represents Public Interface Intelligence, while the larger submerged structure represents deeper forms of AI capability embedded in institutions, operational systems, proprietary knowledge environments, high-authority resource structures, and differentiated rule conditions.

Note: The figure is conceptual and does not imply a linear ranking of intelligence. Depth represents structural position, including access threshold, data environment, institutional embedding, resource authority, real-world intervention capacity, and rule space, rather than intrinsic model capability. “Dark-Domain Intelligence” is used analytically to describe restricted visibility and differentiated operating conditions, not illegality, malicious use, or inherent opacity.

Source: Author’s framework.

3.1 Surface Intelligence Domain

The **Surface Intelligence Domain** refers to the visible, public-facing layer of AI capability. It includes public chatbots, search assistants, office copilots, coding agents, generative media tools, and other consumer-facing AI applications.

This domain is characterized by mass access, high public visibility, open or semi-open interfaces, general-purpose usability, strong product orientation, limited real-world authority, and relatively strong public compliance constraints. Its importance lies in scale: Surface Intelligence shapes public perception, user expectations, consumer adoption, political attention, and the visible market for AI services.

Visibility, however, should not be confused with structural importance. Public-facing systems may be the most visible layer of AI, but they are not necessarily the most consequential. Treating public AI as the whole of AI capability risks overlooking deeper systems embedded in organizations, infrastructures, proprietary knowledge environments, and high-authority decision structures.

3.2 Deep Intelligence Domain

The **Deep Intelligence Domain** refers to the permissioned, professional, organizational, and embedded domain of AI capability. It includes systems integrated into legal analysis, financial modeling, industrial design, biomedical research, logistics, procurement, energy management, enterprise operations, public administration, scientific research, and internal knowledge systems.

This domain is characterized by restricted access, organizational deployment, higher-quality data, workflow integration, professional specialization, operational relevance, feedback loops, and direct productivity impact. Its strategic importance lies in institutional depth: AI is no longer merely an external tool but becomes part of the productive structure of an organization.

Deep Intelligence is likely to become a primary site of AI value accumulation. Its advantage does not come only from stronger models, but from the combination of domain-specific data, expert feedback, institutional memory, workflow integration, and operational continuity. These conditions allow AI capability to become more specialized, more context-aware, and more directly connected to professional output.

3.3 Dark-Domain Intelligence

Dark-Domain Intelligence refers to the restricted-visibility, high-threshold, high-resource, and high-consequence domain of AI capability. It may include systems used in national security, intelligence analysis, military planning, critical infrastructure, emergency governance, strategic resource allocation, high-risk research, or other exceptional environments.

This domain is characterized by small authorized audiences, strong access restrictions, significant resource authority, greater real-world intervention capacity, lower public visibility, higher consequence intensity, and differentiated rule conditions. Its defining feature is not secrecy alone, but the combination of restricted access, concentrated authority, resource connection, real-world consequence, and exceptional operating conditions.

Dark-domain intelligence is not a value judgment. It does not mean that a system is illegal, malicious, opaque, or necessarily more capable than public AI. It describes a structural position within the future distribution of AI power. The same underlying model may have very different implications depending on whether it operates as a public assistant, an enterprise decision tool, or a high-authority system connected to resources, infrastructure, or exceptional rule conditions.

4. Seven Layers of AI Capability

4.1 Layer 1: Public Interface Intelligence

Public Interface Intelligence refers to AI systems that provide general cognitive services to mass audiences through public-facing interfaces, including chat assistants, search assistants, office copilots, coding tools, writing assistants, and generative media systems.

This layer is the visible entrance to AI capability. Its strengths are scale, usability, accessibility, and public adoption; its structural limits are restricted access to proprietary data, limited authority over real-world systems, strong compliance constraints, and limited capacity to command resources or intervene directly in external environments. It shapes public expectations and market behavior, but it should not be mistaken for the full structure of AI power.

4.2 Layer 2: Institutional Embedded Intelligence

Institutional Embedded Intelligence refers to AI systems embedded within the internal knowledge, workflow, and authority structures of organizations, including enterprise knowledge systems, legal and medical workflow tools, engineering design assistants, compliance systems, and research-support platforms.

This layer marks the transformation of AI from an external tool into an institutional capability. The same model can become more valuable when connected to internal documents, professional standards, organizational routines, case histories, decision procedures, and expert review. Its value comes from fit: the alignment between AI capability and institutional knowledge, workflow, and authority.

4.3 Layer 3: Operational Control Intelligence

Operational Control Intelligence refers to AI systems that participate in operational decisions such as resource allocation, process scheduling, infrastructure management, logistics, procurement, maintenance, risk scoring, or system control.

This is the layer where AI begins to affect decisions that change real system states. It may influence where resources are sent, how tasks are scheduled, which risks are prioritized, how inventory is managed, or how infrastructure is operated. Even when humans remain formally in the loop, AI-generated recommendations can become operationally powerful if they structure decisions, prioritize options, or guide resource allocation.

4.4 Layer 4: Proprietary Epistemic Intelligence

Proprietary Epistemic Intelligence refers to AI capability built upon non-public, high-value knowledge materials, expert judgment structures, internal experience, and institutional memory.

Its advantage is epistemic rather than merely computational. It may draw on proprietary research, private transaction data, classified or sensitive records, internal experiments, expert annotations, historical case files, or specialized domain feedback (Bommasani et al., 2021; NIST, 2024).

This layer is not defined by a larger model, but by a more valuable learned world. In professional, commercial, scientific, and strategic settings, the decisive advantage may come from what the AI system is allowed to know. Public AI may possess broad general knowledge, while Proprietary Epistemic Intelligence may possess narrower but more valuable knowledge.

4.5 Layer 5: Objective-Specific Optimized Intelligence

Objective-Specific Optimized Intelligence refers to AI systems optimized around specific objectives, environments, feedback loops, and performance metrics.

This layer does not aim at general knowledge. It aims at task-specific effectiveness in areas such as drug discovery, cyber defense, fraud detection, logistics routing, energy dispatch, industrial process control, or operational planning support (Bommasani et al., 2021; Kapoor et al., 2024).

The strength of this layer is specialization. A narrowly optimized AI system may outperform a more general model within a defined domain because it is trained, evaluated, and improved against specific objectives. Its power comes from alignment between model behavior, task environment, feedback structure, and performance target.

4.6 Layer 6: High-Authority Resource Intelligence

High-Authority Resource Intelligence refers to AI systems controlled by a small number of authorized actors and capable of calling upon significant resources, permissions, and real-world execution interfaces.

This is the layer where AI capability becomes linked to resource command. Relevant resources may include capital, compute, infrastructure, personnel, administrative authority, emergency-response assets, or military capabilities.

The defining issue is not whether the AI system is fully autonomous. The defining issue is whether AI-generated analysis, recommendations, or actions are connected to high-authority resource decisions. Even advisory AI can become highly consequential when it informs actors who control major resources.

4.7 Layer 7: Boundary-Condition Intelligence

Boundary-Condition Intelligence refers to AI systems operating under rule conditions not fully covered by ordinary public, market, platform, or industry rules.

This layer is defined by the rule conditions under which AI systems are authorized to operate. These conditions may include national security settings, military operations, intelligence activities, emergency governance, critical infrastructure, high-risk scientific research, or other exceptional institutional environments (Anderljung et al., 2023; Shevlane et al., 2023).

The same underlying AI capability may produce very different consequences depending on whether it operates under consumer-protection rules, enterprise-compliance rules, military rules, emergency rules, intelligence rules, or other exceptional conditions.

Boundary-Condition Intelligence is therefore the most governance-sensitive layer. It raises questions about authorization, accountability, oversight, escalation control, transparency, and the permissible boundaries of AI-enabled action. Its importance lies not only in what the AI system can do, but in the rule conditions under which it is allowed to act.

5. Why the Layers Are Not a Simple Intelligence Hierarchy

The seven-layer model should not be read as a linear ranking of intelligence. Movement from Public Interface Intelligence to Boundary-Condition Intelligence does not mean that each layer is more intelligent, more advanced, or more capable in every respect than the layer before it.

Each layer reflects a different configuration of capability. Public Interface Intelligence may have the strongest general usability. Institutional Embedded Intelligence may have the strongest organizational fit. Operational Control Intelligence may have the greatest workflow impact. Proprietary Epistemic Intelligence may possess the strongest knowledge advantage. Objective-Specific Optimized Intelligence may achieve the highest task-level performance. High-Authority Resource Intelligence may command the strongest resource access. Boundary-Condition Intelligence may operate under the most differentiated rule conditions.

Depth therefore does not equal intelligence. Depth changes the structure of capability.

This distinction is central to the framework. A public-facing AI model may be stronger in general reasoning, language generation, multimodal interaction, or open-ended usability. A deeper AI system may nevertheless be more consequential because it is connected to proprietary data, operational workflows, institutional authority, resource allocation mechanisms, real-world execution interfaces, or exceptional rule conditions.

AI power is therefore not reducible to model intelligence. It emerges from the interaction between model capability and the system conditions under which that capability is deployed. The relevant question is not only how intelligent an AI system is, but also what it can access, where it is embedded, what resources it can influence, what decisions it can shape, and under what rules it operates.

6. AGI and the Persistence of Stratification

Artificial general intelligence, if achieved, would not erase AI capability stratification.

A highly general AI system may provide broad reasoning, planning, language capability, multimodal understanding, scientific inference, and cross-domain transfer. However, general capability does not automatically confer access to all proprietary data, institutional memory, expert feedback, resource authority, operational interfaces, or special permissions.

AGI may therefore provide a powerful foundation, but specialized AI systems would continue to derive advantage from private data, professional feedback, specific objectives, authority structures, and real-world interfaces. A general system may be broadly capable, while an embedded system may be more useful, more authorized, or more consequential within a specific institutional or operational environment (Bommasani et al., 2021; NIST, 2024).

In this sense, AGI may deepen stratification rather than dissolve it. General AI capability could become a common substrate upon which more specialized, embedded, optimized, and high-authority systems are built. The result would not be a flat AI landscape, but a more complex hierarchy of access, authority, specialization, and rule conditions.

The strategic question is therefore not whether AGI will replace all specialized systems. It is how general AI capability will be distributed, restricted, embedded, optimized, authorized, and governed across different institutional and operational environments.

7. Governance Implications

Governance should follow the structure of AI power. As AI capability becomes stratified across access, data, authority, resources, operational interfaces, and rule space, governance must move beyond public-facing models and account for the system conditions under which AI capability becomes consequential (European Union, 2024; NIST, 2023; OECD, 2019).

Policymakers should distinguish model capability from system capability. Model benchmarks alone cannot capture the practical power of an AI system once it is connected to data, tools, users, institutions, resources, operational interfaces, and decision processes. A moderately capable model embedded in a high-authority environment may be more consequential than a stronger public model operating under strict interface constraints.

Governance should assess authority and intervention capacity. AI systems that can influence resource allocation, infrastructure operation, financial execution, security decisions, administrative action, emergency response, or military planning require different oversight from systems that only generate content for public users. The more directly an AI system can affect real-world outcomes, the stronger the need for authorization controls, auditability, and institutional responsibility.

Governments and institutions should require documentation and auditability for proprietary and embedded AI systems. Full public disclosure may be unrealistic or undesirable where systems rely on proprietary data, sensitive workflows, or security-relevant information. However, limited disclosure does not eliminate the need for governance. Documentation, internal review, sector-specific standards, independent audits, incident reporting, and clear lines of institutional responsibility will become increasingly important (European Union, 2024; NIST, 2024; Office of Management and Budget [OMB], 2025).

Boundary-condition AI requires authorization and escalation controls. Systems operating under emergency, security, intelligence, military, or exceptional regulatory conditions may create heightened risks of opacity, escalation, accountability gaps, and institutional overreach. These systems require explicit authorization rules, review mechanisms, escalation safeguards, and limits on the circumstances under which AI-enabled action may be used (Anderljung et al., 2023; Shevlane et al., 2023).

Finally, policymakers should develop layered governance rather than one-size-fits-all rules. Public-interface AI may require consumer protection, platform accountability, content governance, and user transparency. Deep AI may require sectoral standards, professional liability rules, data governance, audit regimes, and organizational accountability. Dark-domain AI may require authorization controls, red-team evaluation, escalation safeguards, oversight mechanisms, and clear limits on resource-command authority.

The governance problem is therefore not simply how to regulate models, but how to regulate the institutional conditions under which AI capability becomes consequential (NIST, 2023; OECD, 2024).

8. Limitations

This framework is a structural and conceptual model, not a definitive taxonomy of all AI systems. The three domains and seven layers identify recurring patterns in how AI capability is distributed across access, data, authority, resources, operational interfaces, and rule space. In practice, many systems may operate across multiple layers at once, and the boundaries between layers should be understood as analytical rather than absolute.

The framework does not claim that deeper layers are inherently more intelligent, more valuable, more dangerous, or more advanced. Depth refers to structural position, not intrinsic model quality. Public AI may remain more generally capable or more widely useful, while a narrower embedded system may become more consequential because of proprietary data, operational authority, resource access, or differentiated rule conditions.

The framework is also only partially operationalized. It identifies key variables for analyzing the future distribution of AI power, but it does not prescribe universal indicators, weights, thresholds, timelines, or measurement procedures. Nor does it predict that all sectors will follow the same stratification path or that any specific layer will necessarily dominate. Further work is needed to develop empirical metrics, sector-specific applications, governance tests, and comparative case studies.

Conclusion

The future of AI will not be defined only by stronger public models or a single universal AGI market. It is more likely to develop into a stratified capability structure in which different forms of AI power emerge from different combinations of access, data, authority, specialization, resources, operational interfaces, and rule space.

At the visible surface, Public Interface Intelligence will shape mass adoption and social perception. Beneath the surface, Deep Intelligence will accumulate professional productivity through institutional embedding, operational control, proprietary knowledge, and objective-specific optimization. At the deepest levels, Dark-Domain Intelligence will emerge where AI systems operate with high access barriers, concentrated resource authority, significant real-world consequence, and differentiated rule conditions.

The central question for the future is therefore no longer only which AI model is the smartest. It is increasingly who controls what AI can learn, what AI can access, what AI can command, what AI can change, and under what rules AI can operate.

AI Capability Stratification provides a framework for answering that question by shifting analysis from model intelligence alone to the distribution of AI power across access, data, authority, specialization, real-world interfaces, resource command, and rule space.

References

Anderljung, M., Barnhart, J., Korinek, A., Leung, J., O'Keefe, C., Whittlestone, J., Avin, S., Brundage, M., Bullock, J., Cass-Beggs, D., Chang, B., Collins, T., Fist, T., et al. (2023). *Frontier AI regulation: Managing emerging risks to public safety*. arXiv. https://arxiv.org/abs/2307.03718

Bergman, M. K. (2001). The deep web: Surfacing hidden value. *Journal of Electronic Publishing, 7*(1). https://doi.org/10.3998/3336451.0007.104

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021). *On the opportunities and risks of foundation models*. arXiv. https://arxiv.org/abs/2108.07258

European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. *Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Gupta, A., Maynard, S. B., & Ahmad, A. (2021). *The dark web phenomenon: A review and research agenda*. arXiv. https://arxiv.org/abs/2104.07138

Kapoor, S., Stroebl, B., Siegel, Z. S., Nadgir, N., & Narayanan, A. (2024). *AI agents that matter*. arXiv. https://arxiv.org/abs/2407.01502

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

National Institute of Standards and Technology. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence profile* (NIST AI 600-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.600-1

OECD. (2019). *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449). OECD Legal Instruments. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

OECD. (2024). *OECD updates AI Principles to stay abreast of rapid technological developments*. https://www.oecd.org/en/about/news/press-releases/2024/05/oecd-updates-ai-principles-to-stay-abreast-of-rapid-technological-developments.html

Office of Management and Budget. (2025). *Accelerating federal use of AI through innovation, governance, and public trust* (Memorandum M-25-21). Executive Office of the President. https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf

Shevlane, T., Farquhar, S., Garfinkel, B., Phuong, M., Whittlestone, J., Leung, J., Kokotajlo, D., Marchal, N., Anderljung, M., Kolt, N., Ho, L., Siddarth, D., et al. (2023). *Model evaluation for extreme risks*. arXiv. https://arxiv.org/abs/2305.15324

Wu, S. (2026). *Beyond model capability: A system-level framework for AI power* (Policy Brief No. EPINOVA–2026–PB–58). Global AI Governance and Policy Research Center, EPINOVA LLC. https://doi.org/10.67037/epinova.pb.2026.058