

RISE OF THE *KNOWMEBOTS*: PROMOTING THE TWO DIMENSIONS OF AI AGENCY

RICHARD S. WHITT

“We believe that AI will be about individual empowerment and agency at a scale that we’ve never seen before, and that will elevate humanity. . .”¹ - Sam Altman, CEO, OpenAI (2023)

To date, public policy debates over artificial intelligence (AI)—from the EU AI Act, to the Biden Administration’s Executive Order on AI, to the Bletchley Park Declaration—have focused on limiting harms from the largest vertically-integrated Institutional AI providers. Typically, this involves creating greater transparency, oversight, and “guardrails” for the riskiest proposed use cases. While important, this “AI Accountability” agenda standing alone remains incomplete.

Today, no standardized ways, or designated intermediaries, exist for humans—on their own terms—to connect with, query, and gain desired actions from third-party computational systems. Under what could be called an “AI Agency” agenda, however, individuals would use advanced digital technology to actively promote their own best interests. In particular, ordinary people could employ trusted personal AI (PAI) agents to engage directly with the online world. This would include interrogating and contesting consequential decisions made by Institutional AIs and other computational systems. Where AI Accountability focuses on installing guardrails against the riskiest uses, AI Agency creates functional merge lanes in between to enable greater competition, innovation, and consumer choice.

By definition, an agent is able to (1) make decisions and take actions and (2) do so on behalf of someone else. This paper describes two interrelated dimensions of the legal concept of agency: capabilities and relationship. What OpenAI researchers call “agenticity” refers to the ability of an advanced AI system to interact with and make a multitude of decisions in complex environments. This paper introduces the second dimension of “agentiality,” which

1. Sam Altman, CEO, OpenAI, Opening Keynote at OpenAI Developer Day (Nov. 6, 2023).

are forms of relationships that allow the AI to authentically represent its principal. This dimension roughly correlates to the “tetradic” user alignment proposed in a recent paper by Google DeepMind, which posits balancing out relationship values between the user, the AI assistant, the developer, and society. By definition, true agency requires both the advanced capabilities to do things for us (agenticity), and the legitimate relationships to fully represent us (agentiality). This paper’s premise is that both dimensions of agency must be closely aligned if we are to harness advanced computational systems in ways that enhance and promote our human autonomy.

The paper explores one crucial aspect of each agency dimension. First, in order to better promote the advanced capabilities of agenticity—along with greater competition, innovation, and human choice—vertical AI interoperability must provide a technical means for our own PAIs to connect with and influence decisions rendered for us by larger Institutional AIs. The paper relies on a layered framework based on the Levels of Conceptual Interoperability Model (LCIM) as a way to enhance agenticity.

Second, in order to better promote authentic relationships with PAIs, the paper discusses the notion of trusted intermediation, including Net fiduciaries as a design element for the agentiality dimension. Unless acting under established duties of care and loyalty, PAIs risk becoming “double agents”—claiming to represent the principal, while in fact working covertly on behalf of others.

Lastly, the paper proposes the next steps to enact a robust two-dimensional AI Agency agenda, spanning the standards-making, corporate policy, and public policy realms. Elements of these proposed action items are also discussed in the author’s book *Reweaving the Web*, with his proposed trust-based Web overlay called *GliaNet*.

INTRODUCTION: RISE OF THE DIGITAL AGENTS	53
I. THE ROAD TO “KNOWMEBOTS”	54
II. NEW AI ECOSYSTEMS	56
A. <i>The Rise of Institutional AIs and Their Ecosystems</i>	56
B. <i>Here Come the Digital Agents</i>	60
III. CREATING AN AUTHENTIC AGENT – IN TWO	
DIMENSIONS	62
A. <i>Dimension One: Agenticity</i>	62
B. <i>Dimension Two: Agentiality</i>	63
C. <i>Two Dimensions of Agency, Together</i>	65
IV. AGENTICITY AND AI INTEROP	66

A.	<i>Benefits: Innovation, Competition, Human Agency</i>	67
B.	<i>Unpacking Digital Forms of Interop</i>	69
C.	<i>Introducing Vertical AI Interop</i>	71
D.	<i>Other Related Edgetech: SLMs and Middleware</i>	75
1.	Small Language Models	75
2.	Middleware.....	75
V.	AGENTIALITY AND TRUSTWORTHY INTERMEDIATION	77
A.	<i>A Pathway to Fiduciary AI Agents</i>	80
B.	<i>The GliaNet Initiative’s PEP Model</i>	83
VI.	ROBUST AGENCY: BETTER TOGETHER.....	86
VII.	THE POLICY GAP	91
A.	<i>AI Accountability Agenda: Helpful, but Insufficient</i>	91
B.	<i>Human Agency via AI Agenda</i>	92
VIII.	SOME NEXT STEPS	93
A.	<i>The Role of Software Standards</i>	94
1.	Open Interop APIs: A Sufficient Starting Point?	94
2.	Open Standards For Digital Agents	95
B.	<i>Relying on Market Adoption</i>	96
1.	Voluntary Interop?.....	96
2.	Voluntary Association of the Willing?	97
C.	<i>Public Policy: Codifying Our Rights and Their Duties</i>	98
1.	The Current Landscape for Interop	98
2.	The Current Landscape for Agential Relationships	
	100
IX.	A PROPOSED RIGHTS/DUTIES FRAMEWORK.....	101
	CONCLUSION	103

INTRODUCTION: RISE OF THE DIGITAL AGENTS

2023 was the year of Generative AI, creating new forms of content by using large language models (LLMs). But what is in store for us in 2024 and beyond? The technologists and business pundits seem to agree: We are witnessing the dawn of the era of personal AI systems. From voice assistants on phones, to automated shopping lists, to predictive text in emails, these systems are becoming pervasive in everyday life. Additionally, today’s promise of individualized digital bots is morphing from information-only search retrievers, to multi-tasking virtual assistants, to full-fledged agents that understand their humans at deep levels and take actions (ostensibly) on their behalf. How then should we be thinking about these new personalized instantiations of advanced AI technology? This paper explores the two interrelated definitional aspects of agency: capabilities to do things for us and relationships to fully represent us. As shown, achieving a more authentic personal AI (PAI) agent will require (1) interoperability

between AI systems to optimize advanced capabilities, and (2) trustworthy intermediation to optimize alignment of digital relationships.

I. THE ROAD TO “KNOWMEBOTS”

In March 1988, Bob Kahn and Vint Cerf—actual “co-fathers of the Internet”—released a draft paper called “An Open Architecture for a Digital Library System and a Plan for its Development.”² Behind its innocuous-sounding title, the document explored the concept of an open architecture for national information infrastructure, which its authors called the “Digital Library System” (DLS). Intriguingly, the architectural framework included mobile software agents called “Knowbots,” which were tasked with performing various mediating services for their end users:

Knowbots ... are active intelligent programs capable of exchanging messages with each other and moving from one system to another in carrying out the wishes of the user. They may carry intermediate results, search plans and criteria, output format and organization requirements and other information relevant to the satisfaction of a user’s query. A Knowbot is typically constructed on behalf of a user at his Personal Library System and dispatched to a Database Server which interfaces the network to a particular database or set of databases.... In the future, we expect to witness the development of databases systems [*sic*] with built-in mechanisms for housing and catering to resident or transient Knowbots.³

The Kahn/Cerf paper goes on to describe a class of trusted Knowbots called “couriers.” These “couriers” have a special responsibility to look after selected objects, such as documents or databases, on behalf of their rights holders.⁴

The Knowbot’s five key aspects are that it (1) is constructed on behalf of the end user, (2) responds to the user’s queries, (3) moves seamlessly from one information system to another, (4) interacts in various ways with those larger systems, and (5) guards the sensitive content of its end user.

2. ROBERT E. KAHN & VINTON G. CERF, AN OPEN ARCHITECTURE FOR A DIGITAL LIBRARY SYSTEM AND A PLAN FOR ITS DEVELOPMENT (CORP. FOR NAT’L RSCH. INITIATIVES, 1988).

3. *Id.* at 17.

4. *Id.* at 34.

Fast forward to 35 years after the Knowbots paper. A Silicon Valley audience member asked Vint Cerf about the sudden advent of generative AI.⁵ Cerf's advice? Don't rush into investing in these businesses. As he explained:

There's an ethical issue here that I hope some of you will consider. Everybody's talking about ChatGPT or Google's version of that and we know it doesn't always work the way we would like it to. Be thoughtful.... You were right that we can't always predict what's going to happen with these technologies and, to be honest with you, most of the problem is people — that's why we people haven't changed in the last 400 years, let alone the last 4,000. They will seek to do that which is their benefit and not yours. So, we have to remember that and be thoughtful about how we use these technologies.⁶

Looking beyond 2024, technological progress in artificial intelligence is only accelerating. Generative AI has captured the world's headlines with the promise of vast computational resources at everyone's fingertips. Digital agents are evolving beyond the Kahn/Cerf conception of mere knowledge retrievers to encompass complex forms of autonomy and agency on behalf of ordinary people. In this regard, one could say that we are moving rapidly from a world of Kahn and Cerf's more linear Knowbot to a more dynamic "KnowMeBot." The ramifications for this shift, combining the informational elements of the Web and the behavioral elements of the individual, are immense.

Indeed, one can posit that the shift towards digital agents is a continuation of the evolving and overlapping ways people are able to interact with the Web. From about 1995 to 2010, the browser was the predominant interface provided by online companies to access websites. Heading into the 2010s, with the introduction of the iPhone and Android-based mobile devices, the embedded Web application became another such interface. As AI apps populated our mobile and fixed devices, voice and gesture commands were made available to end users. Now, in 2025, the digital agent is

5. See Jennifer Elias, *Father of internet warns: Don't rush into A.I. just because ChatGPT is really cool*, CNBC (Feb. 14, 2023, 9:02 AM), <https://www.cnbc.com/2023/02/14/father-of-the-internet-warns-dont-rush-investments-into-chat-ai.html> [<https://perma.cc/B864-ZKPQ>]; see also Prarthana Prakash, *Father of the internet' Vint Cerf tells investors to think before pouring money into A.I. bots like ChatGPT: 'There's an ethical issue'*, FORTUNE MEDIA (Feb. 15, 2023, 5:22 AM), <https://fortune.com/2023/02/15/father-internet-vint-cerf-warns-investors-ai-bots-chatgpt-ethical-issue> [<https://perma.cc/L569-E8HR>].

6. Elias, *supra* note 5; Prakash, *supra* note 5.

poised to become the leading interface between humans and the Web, encompassing all of these and other mediation points. At this juncture, we now have the opportunity to step back and assess where we are, and (more critically) where the larger Web/AI platforms want to take us.

This paper takes up Cerf's injunction to "be thoughtful" and carefully consider the ethical issues that accompany our most advanced technologies. Particularly, as computational systems operated by corporations and governments continue to permeate our lives, society must consider where our own human agency can still coexist, let alone flourish, with these systems. So long as the people behind the scenes of these Institutional AIs will do "that which is their benefit and not yours," our personal and collective interests will never be fully realized.

One proffered technological answer is the digital agent, which can be positioned by platform companies as an AI-based assistant to represent us. But who exactly should own and operate these "KnowMeBots?" Should it be those same institutions or ourselves? And if some of us prefer to inhabit our own digital destinies, how exactly can we get there from here?

This paper introduces the two dimensions of AI agency: "agenticity," or capabilities, and "agentiality," or relationship. To better promote advanced capabilities—along with greater competition, innovation, and human choice—the paper first discusses the notion of "vertical AI interop," a technical means for our own personal digital agents to connect with and influence the decisions rendered for us by the larger Institutional AIs. To better promote authentic relationships, the paper discusses the notion of trusted intermediation, including digital fiduciaries as the optimal way to govern these technologies. Consistent with the author's Giant initiative,⁷ the premise is that functional openness tools—like vertical AI interop—and corporate governance implements—like Net fiduciaries—are crucial to fully harness advanced computational systems in ways that enhance and promote our human autonomy and agency.

II. NEW AI ECOSYSTEMS

A. *The Rise of Institutional AIs and Their Ecosystems*

Modern computational systems contain three elements: (1) lots of data, (2) advanced algorithms, and (3) software-derived interfaces. In turn, such systems being created, trained, and

7. See RICHARD WHITT, REWEAVING THE WEB (The Book Shelf Ltd., 2024).

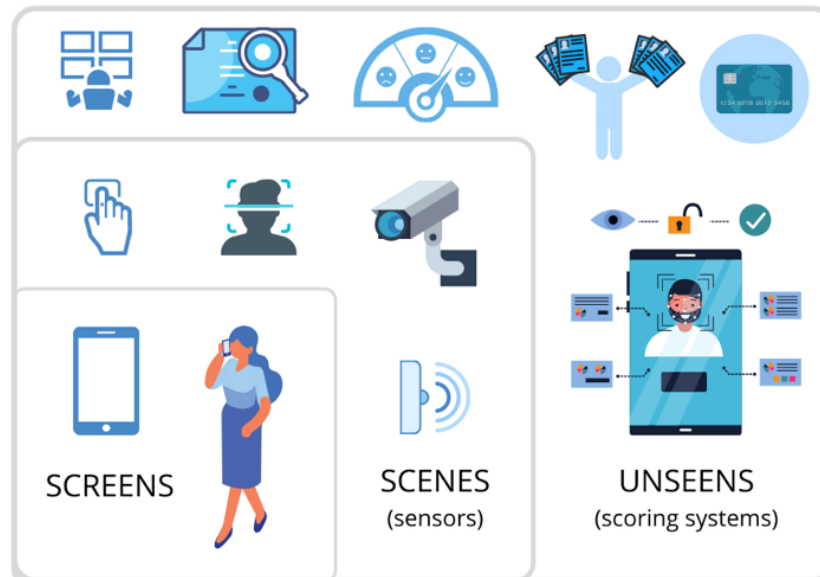
deployed by large incumbent corporations and governments can be thought of as “Institutional AIs.” These systems churn through a person’s data to discern actionable insights. They can be found via interfaces in three distinct environments: (1) our connected screens, (2) our environmental scenes, and (3) bureaucratic unseens.⁸

Online screens lead us to search engines, social media platforms, and countless other Web portals in our lives. Institutional AIs and the computational systems behind them render recommendation engines that guide us to places to shop, videos to watch, and news content to read. More ominously, these systems (with their user engagement imperative) can potentially prioritize the delivery of “fake news,” extremist videos, and dubious advertising.

Environmental scenes are the “smart” devices—cameras, speakers, microphones, sensors, beacons, actuators—scattered throughout our homes, offices, streets, and neighborhoods. These computational systems gather from these interfaces a mix of personal (human) and environmental (rest of world) data. The Ring doorbell placed by your neighbor across the street is but one example.

Bureaucratic unseens are hidden behind the walls of governments and companies. These computational systems render judgments about our basic necessities and personal interests. These decisions—often utilizing Institutional AIs—can include life-altering situations, such as who gets a job or who gets fired, who is granted or denied a loan, who receives what form of healthcare, and who warrants a prison sentence. Unlike with screens, we don’t get the benefit of an actual interface as a mediation point.

8. Richard Whitt, *Hacking the SEAMs: Elevating Digital Autonomy and Agency for Humans*, 19.1 COLO. TECH. L. J. 135, 144–45 (2021).



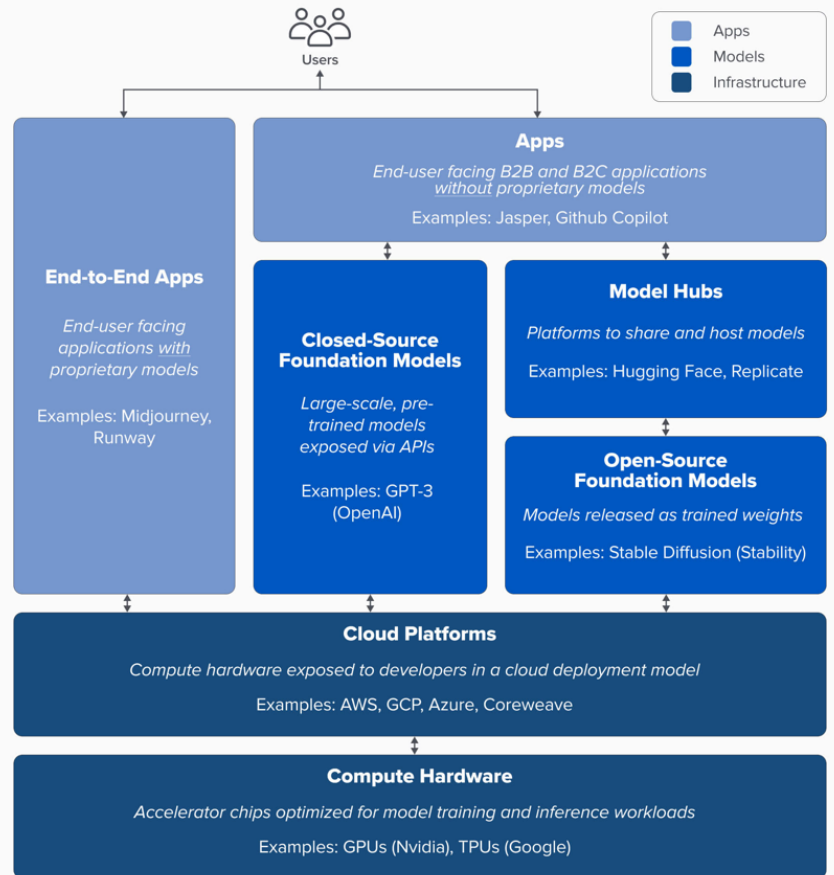
As such, Institutional AIs pose a threefold societal challenge: they are *pervasive* (operating behind interfaces fronting as online screens, environmental scenes, and bureaucratic unseens), *consequential* (operating decision engines from online recommendation systems, to speech bots in our living rooms, to analytics networks in every corporate and government office), and *inscrutable* (utilizing virtual “black boxes” that even experts often cannot understand).⁹ While the sheer power and reach of these cutting-edge technologies is impressive, there usually is only a small handful of entities truly calling the shots: the “developers” and “deployers” of these systems.

While there is no definitive assessment of the new generative AI ecosystem being created from the various elements of computational systems, Andreesen Horowitz has put together a preliminary sketch.¹⁰

9. *Id.* at 190.

10. MATT BORNSTEIN ET AL., *Preliminary generative AI tech stack* (illustration), in *Who Owns the Generative AI Platform?*, ANDREESSEN HOROWITZ (Jan. 19, 2023), <https://a16z.com/who-owns-the-generative-ai-platform> [<https://perma.cc/5VFF-MMLW>].

Preliminary generative AI tech stack



aló Enterprise

As the authors note, “this is not a market map, but a framework to analyze the market.”¹¹ The paper points out how the generative AI stack can be divided into three layers:

- **Applications:** integrate generative AI models into a user-facing product, either running their own applications (“end-to-end apps”) or relying on a third-party public or proprietary application programming interface (API).
- **Models:** enable AI products, using either as public or proprietary APIs.

11. *Id.*

- **Infrastructure:** providing cloud platforms and hardware that run workloads for generative AI models.¹²

For our purposes, it is enough to observe that the infrastructure and model layers are largely represented by a few vertically-integrated companies: Google, Microsoft, Meta, Amazon, and Apple. Even when other entities are involved, such as OpenAI and Anthropic, they too benefit from investments by these major players. Thus, we are left with a small number of Institutional AIs that bring their substantial preexisting market position from the Web.

B. *Here Come the Digital Agents*

With the rise of “first-generation” virtual assistants, such as Alexa, Siri, and Google Assistant, consumers have already been purchasing mobile and home devices that include AI-based bots. Moving into 2025, however, the buzzword has been the chatbot. Built on generative AI platforms, “second-generation” chatbots are poised to revolutionize many aspects of modern life, from education to employment to healthcare, and more.

Tech leaders have not overlooked these recent developments. Sam Altman has said that what he calls “personal agencies” are “going to be a giant thing of the next couple of years,” that “a lot of folks are gonna want.”¹³ Bill Gates has opined that these digital agents will constitute a “revolution,” and “utterly change how we live our lives, online and off.”¹⁴ As he sums up, “in short, agents will be able to help with virtually any activity and any area of life. The ramifications for the software business and for society will be profound.”¹⁵

This incoming wave of personal digital agents—also known as Personal AIs (“PAIs”)—could be thought of as “KnowMeBots,” all of which claim to serve the individual. OpenAI has introduced GPTs to undertake certain tasks for end users of ChatGPT-4.¹⁶ Google has brought out various forms of Gemini as the “AI collaborator” front-

12. *Id.*

13. What Now? With Trevor Noah, *Sam Altman Speaks Out About What Happened at OpenAI*, LISTEN NOTES, at 43:42 (Dec. 7, 2023).

14. Bill Gates, *AI is about to completely change how you use computers*, GATES NOTES (Nov. 9, 2023), <https://www.gatesnotes.com/AI-agents> [<https://perma.cc/4L6B-38HT>].

15. *Id.*

16. *See Introducing GPTs*, OPENAI, <https://openai.com/blog/introducing-gpts> [<https://perma.cc/SC4C-A3QY>] (Nov. 6, 2023).

end to its AI platform,¹⁷ along with “Pixie” as a digital assistant embedded on its Pixel phones.¹⁸ Microsoft has introduced Copilot, “your everyday AI companion,” as part of its Office 365 ecosystem.¹⁹ As Microsoft puts it, “soon there will be a Copilot for everyone and for everything you do.”²⁰ Elon Musk’s X is introducing Grok AI,²¹ a chatbot with direct access to the X platform, while Anthropic has its own generative AI-powered agent, Claude.²² Amazon’s Q is an assistant designed for office tasks in business environments.²³ Inflection, a company funded by Microsoft and Eric Schmidt, has launched “Pi,” a chatbot designed to “project an empathetic and caring persona” using the startup’s proprietary LLM.²⁴ These are just a few examples of what could be thought of as Institutional PAIs, and the list goes on.

As Bruce Schneier explains, these Institutional PAIs serve as “double agents” because they appear to be operating under our control, as “our” PAIs.²⁵ He also notes that they will seem more relational and intimate to us, even as their companies attempt to anthropomorphize them.²⁶ Even the interfaces themselves will be more “personal.” With the introduction of wearable devices, such as smart glasses and pins, and eventually even bodily implants, the

17. See *Gemini*, GOOGLE, <https://bard.google.com> [<https://perma.cc/4QND-ZY83>] (last visited Nov. 7, 2024).

18. *Meet ‘Pixie,’ the Gemini-powered AI assistant Google is developing for Pixel phones*, THE ECONOMIC TIMES ONLINE (Dec. 15, 2023, 2:57 PM), <https://economictimes.indiatimes.com/magazines/panache/meet-pixie-the-gemini-powered-ai-assistant-google-is-developing-for-pixel-phones/articleshow/106018446.cms?from=mdr> [<https://perma.cc/3T78-ALMV>].

19. *Copilot*, MICROSOFT, <https://copilot.microsoft.com> [<https://perma.cc/Q5HU-YPDZ>] (last visited Nov. 7, 2024).

20. Colette Stallbaumer, *Introducing Microsoft Copilot Studio and new features in Copilot for Microsoft 365* (Nov. 15, 2023), <https://www.microsoft.com/en-us/microsoft-365/blog/2023/11/15/introducing-microsoft-copilot-studio-and-new-features-in-copilot-for-microsoft-365> [<https://perma.cc/8V2A-M7BZ>].

21. Elon Musk (@elonmusk), X (Nov. 4, 2023, 10:24 AM), <https://twitter.com/elonmusk/status/1720839331365929290?lang=en> [<https://perma.cc/YN2R-YKB6>].

22. *Introducing Claude*, ANTHROPIC (Mar. 14, 2023), <https://www.anthropic.com/index/introducing-claude> [<https://perma.cc/HQP8-RACH>].

23. *Amazon Q*, AMAZON, <https://aws.amazon.com/q> [<https://perma.cc/P945-EVAR>] (last visited Nov. 7, 2024).

24. Eric H. Schwartz, *‘Personal’ Generative AI Startup Inflection AI Raises \$1.3B From Microsoft, Nvidia, Bill Gates, and Eric Schmidt*, VOICEBOT.AI (June 29, 2023, 2:00 PM), <https://voicebot.ai/2023/06/29/personal-generative-ai-startup-inflection-ai-raises-1-3b-from-microsoft-nvidia-bill-gates-eric-schmidt> [<https://perma.cc/H4G2-Q47Z>].

25. Bruce Schneier, *AI and Trust*, SCHNEIER ON SECURITY (Dec. 4, 2023, 7:05 AM), <https://www.schneier.com/blog/archives/2023/12/ai-and-trust.html> [<https://perma.cc/ADL3-5VYV>].

26. *Id.*

illusion of human ownership and control over these agents becomes ever stronger.

However, with the ongoing addition of human-Web interfaces along the continuum from online screens to environmental scenes to cloud-based “unseens,” these PAIs are poised to become our predominant way of mediating with the Web. The question then arises: for whom exactly do these digital agents actually work?

III. CREATING AN AUTHENTIC AGENT – IN TWO DIMENSIONS

By definition,²⁷ there are two interrelated dimensions to being an agent: (1) making decisions and taking actions and (2) making and taking said actions on behalf of someone else. In the context of a technology, such as a PAI or other forms of a digital agent, these intrinsic elements can be thought of as “agenticity” (functional capability to behave like an agent, by performing or undertaking certain activities)²⁸ and “agentiality” (authorized representation to act for another person).²⁹ We will explore both dimensions below.

A. *Dimension One: Agenticity*

In December 2023, OpenAI released a white paper discussing what it calls the “agenticness” or “agenticity” of personal AI assistants.³⁰ The white paper defines an AI system’s “agenticness” as the “degree to which it can adaptably achieve complex goals in complex environments with limited direct supervision.”³¹ One can also think of this as “environmental alignment.”

The paper lists four components of agenticness: (1) goal complexity, (2) environmental complexity, (3) adaptability, and (4) independent execution.³² Goal complexity refers to the range and depth of responses, while environmental complexity reflects

27. See, e.g., *Agent*, MERRIAM-WEBSTER, <https://www.merriam-webster.com/dictionary/agent> [<https://perma.cc/DS89-D4EE>] (last visited Nov. 17, 2024); *Agency*, CORNELL LEGAL INFORMATION INSTITUTE, <https://www.law.cornell.edu/wex/agency> [<https://perma.cc/MZ5M-3EQV>] (last visited Nov. 17, 2024).

28. *Agentic*, WIKTIONARY, <https://en.wiktionary.org/wiki/agentic> [<https://perma.cc/34VU-YB9E>] (last visited Sept. 2, 2024).

29. One can imagine as a matter of symmetry adding a third dimension that covers the principal individual herself, but we can leave that scenario to another day.

30. YONADAV SHAVIT ET AL., OPENAI, PRACTICES FOR GOVERNING AGENTIC AI SYSTEMS (2023).

31. *Id.* at 4.

32. *Id.* (where for each component, the paper posits various ways of measuring how well an AI system can meet its goals in uncertain or complex situations).

achievements across domains, stakeholders, and time horizons.³³ Adaptability is the ability to react to novel or unexpected circumstances.³⁴ Independent execution requires limited degrees of human intervention or supervision.³⁵ Those systems exhibiting overall high degrees of agentiveness are referred to as “agentic AI systems.”³⁶

The paper also lists seven outstanding challenges requiring additional attention: (1) evaluating suitability for the task, (2) circumstances for human approval, (3) setting default behaviors, (4) legibility of actions, (5) monitoring, (6) attributability, and (7) maintaining control.³⁷ Many of these practices amount to ensuring that the user/agent relationship adequately meets the developer’s expectations.

The OpenAI paper posits the desirability of examining the range of effects and best practices that can become relevant as a system’s agentiveness increases.³⁸ This includes allocating accountability for harms from such systems.³⁹ The paper also discusses the second of the seven challenges, “constraining the action-space and requiring approval,” ensuring that there is a “human in the loop” to limit egregious harms.⁴⁰

Open questions include how a user or system deployer determines which interactions require human approval and how such approval should be gained.⁴¹ However, the paper does not directly address what could be thought of as the second dimension of agency: valid authorization for an AI system to act on behalf of a person.

B. Dimension Two: Agentivity

In contrast with OpenAI’s more functional concept of agentivity, what could be termed an AI system’s “agentivity” adheres more closely with the second dimension laid out above: the degree to which the AI is legally/formally/ethically authorized to represent the principal human being. The OpenAI white paper uses the popular computer term “user-alignment” and defines it as “the

33. *Id.*

34. *Id.*

35. *Id.*

36. *Id.* at 5.

37. *Id.* at 7-15.

38. *Id.* at 5.

39. *Id.* at 3, 7.

40. *Id.* at 9.

41. *Id.* at 10.

propensity of an AI model or system to follow the goals specified by a user.”⁴² In terms of that work, it seems like a reasonable definition for agentiality. But that is not quite the last word on the topic.

Interestingly, on the heels of the OpenAI white paper, a large group of researchers aligned with Google DeepMind issued its own extensive paper focused on AI. This April 2024 paper, entitled “The Ethics of Advanced AI Assistants,”⁴³ can be seen as picking up where OpenAI left off by helping answer the question of user alignment. The paper posits that “an AI assistant is aligned with a user when it benefits the user, when they ask to be benefited, in the ways they expect to be benefited.”⁴⁴

The Google DeepMind paper acknowledges that putting the user/AI assistant relationship on a sound ethical footing is a necessary task, but not a sufficient one given the wider social context.⁴⁵ Instead, the paper proposes that value alignment should encompass a “tetradic relationship” between the user, the AI assistant, the developer, and society.⁴⁶ We will explore some aspects of this proposed tetradic relationship as part of our larger consideration of the second agency dimension of agentiality. The larger thesis is that any searching inquiry into agenticity is incomplete without a similar exploration of agentiality. Indeed, without an adequate confirmed degree of agentiality, any actions taken by the AI system “on behalf of” a person should be considered unwarranted—if not downright unethical.

42. *Id.* at 2 n.4.

43. Iason Gabriel, Arianna Manzini & Geoff Keeling et al., *The Ethics of Advanced AI Assistant*, GOOGLE DEEPMIND (Apr. 19, 2024), <https://arxiv.org/abs/2404.16244> [<https://perma.cc/9DW8-85EF>].

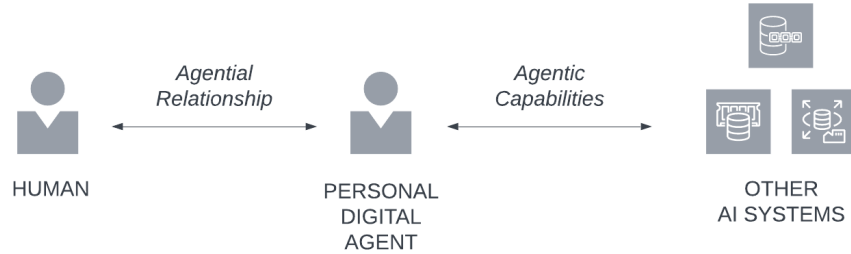
44. *Id.* at 34.

45. *Id.* at 41-42.

46. *Id.* at 42-43. It should be noted that much of the paper’s discussion of human ethics relies on conceptions typically found in Western-oriented philosophy and science. *See, e.g., id.* at 45-49. As research continues into developing and applying human/AI ethical frameworks, participants should be mindful that there are other ways of conceiving human flourishing. One such example is the ubu-Ntu philosophy more prevalent in Africa. *See*, Sabelo Mhlambi, *From Rationality to Relationality: Ubuntu as an Ethical and Human Rights Framework for Artificial Intelligence Governance*, CARR CTR. FOR HUM. RTS. POL’Y, HARV. KENNEDY SCH., Spring 2020 (arguing rational personhood as the basis for mechanical thinking and AI systems fails to address the more relational, contextual, and interconnected world view provided by the ubu-Ntu philosophy). Hopefully the comprehensive research project proposed in the Google DeepMind paper will encompass consideration of AI ethics grounded in these “alternative” ethics systems.

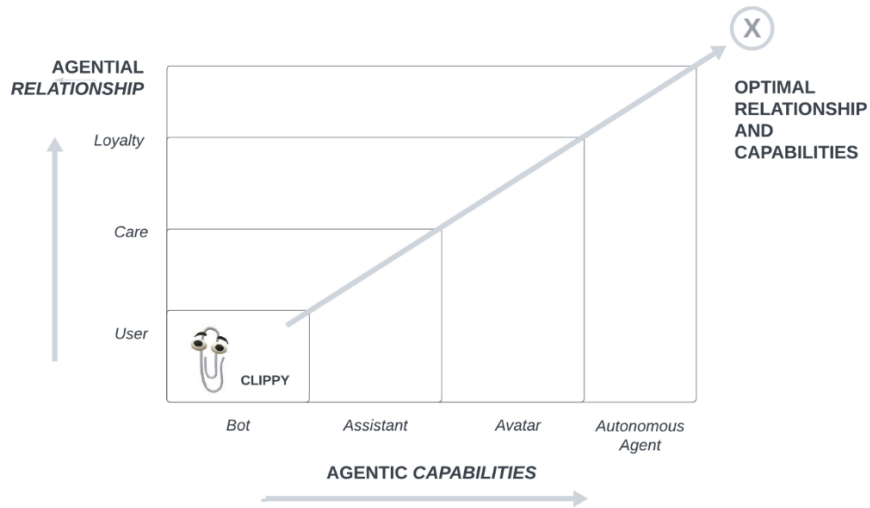
C. Two Dimensions of Agency, Together

Here is a proposed schema for a PAI that combines the outward-facing capabilities with the inward-facing relationship:



In the AI context, one can posit that the greater the capabilities for environmental or situational alignment with other systems, the greater the need for end user alignment. Alternatively, the greater the degree of observable agentiality—the deeper the relationship—the larger the valid action space available for agenticity. This means there should be acknowledged proportional tradeoffs between agenticity (technical capabilities) and agentiality (authorized representation). Per Spider-Man’s famous credo, “With great power, comes great responsibility.”

Here is one way to capture that crucial interrelation between the two dimensions of AI agency:



In everyday terms, this means that the more consensual the agreement to certain forms of AI-assisted decision-making, the more advanced kinds of actions that AI system can perform in the world.

Below we will explore how both dimensions can be optimized by way of technology and governance implements. In brief, greater agenticity can be achieved through the addition of vertical AI interop, while greater agentiality can be achieved through the addition of trustworthy intermediation.

IV. AGENTICITY AND AI INTEROP

Agenticity is about performing tasks for someone; the wider the possible canvas, the more opportunities exist for such actions. Today, a PAI is technologically limited to operating on its sponsoring platform. Without access to the computational resources residing on other platforms, a PAI is unable to carry out a host of potential functions related to those systems. The challenge then is finding ways to extend a PAI's capabilities to other AI systems. The proffered solution is AI interoperability.

Interoperability is defined as the ability of heterogeneous networks to connect and communicate seamlessly with one another.⁴⁷ For example, to become part of the Internet, operators of individual networks voluntarily connect to other networks. The IETF's Request for Comment (RFC) 1958 puts it plainly: the "why" of the Internet is connectivity, moving data packets seamlessly from Point A to Point B.⁴⁸ Such open connectivity requires technical cooperation between different network providers. This connective tissue, combined with open standards, modularity, and the agnostic IP protocol, is responsible for the incredible generative power of the Internet.⁴⁹

Technological interoperability long precedes the Internet, with the telephone, telegraph, and radio systems all relying on interoperable protocols. "The concept even predates electrical

47. *Interoperability*, WIKIPEDIA, <https://en.wikipedia.org/wiki/Interoperability> [<https://perma.cc/UGL4-8RUV>] (as of Nov. 17, 2024).

48. INTERNET ENG'R TASK FORCE, *Request for Comment 1958* (B. Carpender ed., June 1996), at 2. To Kevin Werbach, "the defining characteristic of the Net is ... a relentless commitment to interconnectivity." Kevin Werbach, *Only Connect*, 22 BERKELEY TECH. L. J. 1233, 1273 (2007).

49. See Richard Whitt, *Hiding in the Open: How Tech Network Policies Can Inform Openness by Design (and Vice Versa)*, 3 GEO. L. TECH. REV. 28, 38-45 (2018).

communications, and can be seen in the development of everything from railroad gauges to bullet calibers.”⁵⁰

Following these impressive legacies, the next logical step is creating interoperability for disparate AI systems. If AI fulfills the promise of serving as the crucial platform of the 21st century, we need to take a thoughtful look at the interop fabric that can bind us together – if it is actually adopted and utilized.

A. *Benefits: Innovation, Competition, Human Agency*

In the book *Interop*, Palfrey and Gasser observe that “[t]he benefits and costs of interoperability are most apparent when technologies work together so that the data they exchange prove useful at the other end of the transaction.”⁵¹ Without interoperability at the lower layers of the Internet, interoperability at the higher human and institutional layers is often impossible.⁵² Palfrey and Gasser’s concept of “interop” is to “embrace certain kinds of diversity not by making systems, applications, and components the same but by enabling them to work together.”⁵³ Thus, if the underlying platforms are open and designed with interop in mind, then all players—including end users and intermediaries—can contribute to the development of new products and services.⁵⁴

Other important voices in the online technological field agree. For example, in March 2020, New America published a detailed research report citing the benefits of interoperability in the context of online platforms. The authors found that interop has “a unique ability to promote and incentivize competition—especially competition between platforms,” as well as offering end users “greater privacy and better control over their personal data generally.”⁵⁵ In particular, where interop gives users more choices over what entities have access to their data, they can utilize more secure data protection environments.⁵⁶

50. Becky Chao & Ross Schulman, *Promoting Platform Interoperability*, NEW AMERICA FOUNDATION (May 13, 2020), <https://www.newamerica.org/oti/reports/promoting-platform-interoperability> [https://perma.cc/WLF4-YV25].

51. JOHN PALFREY & URS GASSER, *INTEROP: THE PROMISE AND PERILS OF HIGHLY INTERCONNECTED SYSTEMS* 22-23 (2012).

52. *Id.* at 23.

53. *Id.* at 108.

54. *Id.* at 121.

55. Chao & Schulman, *supra* note 50, at 6.

56. *Id.* at 19.

Similarly, the U.S. Federal Trade Commission (FTC) has also highlighted the important benefits derived from interoperability:

Indeed, there are aspects of technology that people may take for granted while navigating their daily routines that turn on interoperability. Web pages can display regardless of web browser. Emails can be sent and received regardless of email provider. Computer accessories, including keyboards, mice, and monitors, can be plugged into most computers regardless of manufacturer. Interoperability can also enhance consumer choice and facilitate switching between products, and thereby enhance competition.⁵⁷

Web platform companies have also acknowledged the importance of interop as a way for users to share access to their data. The Data Transfer Project (DTP) founded by Google, Apple, and Facebook – now called the Data Transfer Initiative -- has been building an open-source platform for moving data between various platforms.⁵⁸ As the sponsors put it: “DTI partners help translate principle to practice, catalyzing greater user agency and empowerment by committing dedicated policy and engineering resources to the promotion of data portability.”⁵⁹

Most recently, Cory Doctorow dedicated an entire book to the role of interoperability in curtailing the Web’s pervasive “winner takes all” network effects.⁶⁰ Doctorow explains that “interoperators” who plug new technologies into existing platforms and services help lower switching costs between providers and allows users to set the terms for using technologies.⁶¹ He concludes that interop brings “immediate, profound relief” from “manipulation, high-handed moderation, surveillance, price-gouging, disgusting or misleading algorithmic suggestions ... the

57. Staff in the Off. of Tech. and the Bureau of Competition, *Interoperability, Privacy, and Security*, F.T.C. (Dec. 21, 2023), <https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/12/interoperability-privacy-security> [<https://perma.cc/XZ5P-LY8F>].

58. *Data Transfer Project*, WIKIPEDIA, https://en.wikipedia.org/wiki/Data_Transfer_Project [<https://perma.cc/FD3R-QTC9>] (as of Nov. 17, 2024).

59. DATA TRANSFER INITIATIVE, <https://dtinit.org> [<https://perma.cc/8H37-BDRK>] (last visited Sept. 9, 2024).

60. CORY DOCTOROW, *THE INTERNET CON: HOW TO SEIZE THE MEANS OF COMPUTATION* (2023). Doctorow lauded “competitive compatibility,” where upstart companies build innovative new products on top of preexisting ones. Cory Doctorow, *Competitive Compatibility: Let’s Fix the Internet, Not the Tech Giants*, COMMC’N OF THE ACM (Oct. 1, 2021), <https://cacm.acm.org/magazines/2021/10/255710-competitive-compatibility/abstract> [<https://perma.cc/GH3Q-BVMB>].

61. *Id.* at 3.

whole panoply of technology’s sins.”⁶² In somewhat less hyperbolic terms, it is sufficient to observe that in broad strokes, interoperability can increase competitive opportunities, foster innovation, and provide consumers with real choice and control over their technologies.

B. *Unpacking Digital Forms of Interop*

In March 2022, the Centre on Regulation in Europe (CERRE) published a paper detailing the various aspects of interop, including pros and cons for its potential use cases in the digital economy.⁶³ The CERRE report distinguishes between horizontal interops, where the products or services operate at the same level of the value chain, and vertical interops, where they operate at different levels. An example of horizontal interop is email, where Google’s Gmail and Microsoft’s Outlook services compete in the same market but also allow users to connect with each other. An example of vertical interop is online game developers having the ability to sell their apps in different Web application stores. With horizontal interop, the network effects enjoyed by a single platform or service are aggregated into market-wide network effects that become a public good.⁶⁴ The paper concludes that “vertical interoperability is indeed a powerful instrument for regulating digital bottlenecks.”⁶⁵ Nonetheless, there are a host of crucial aspects that should be considered.

The CERRE report lays out three levels of technical integration that apply to any form of interop: (1) *Protocol*, where products and services can interconnect and work together; (2) *Data*, where data can be exchanged in real-time between different services, using an open API; and (3) *Full protocol*, where substitute services can interoperate.⁶⁶ This last version enables a firm to offer products and services that can access a competitor’s user base, thereby sharing in network effects.⁶⁷

In turn, vertical interoperability is closely connected to the concept of modularity—one of the foundational design principles of the Internet. It allows end users to mix and match system components, facilitate design innovations, strengthen competition

62. *Id.*

63. MARC BOURREAU ET AL., INTEROPERABILITY IN DIGITAL MARKETS, CTR. ON REGUL. IN EUR. (2022).

64. *Id.* at 19.

65. *Id.* at 45.

66. *Id.* at 13.

67. *Id.*

in complementary markets, and increase the overall value of the platform or service. On the other hand, vertical interop can make it more difficult for a platform provider to capture the full value of the platform because multiple competing applications are sharing access to the platform as a necessary resource. Beyond such challenges, many commenters still conclude that vertical interop is generally pro-competitive.⁶⁸

Vertical interop can occur either within a platform, or cross-platform. The *within-platform* version, also called vertical compatibility, allows third party developers to supply complements for a given product, service, or platform. This typically involves the provision of open (public) APIs, where the platform can restrict access or require permission. *Cross-platform* interop is the more powerful of the two, which builds on the within-platform version by allowing third party developers to offer their complementary products or services to different platforms. This means the interfaces, including APIs, must be standardized to some extent. Thus, cross-platform interop promotes competition between digital platforms or ecosystems.⁶⁹

Cross-platform interop also allows for “equitable interoperability,” which one set of authors defines as the idea “that an entrant can not only join the platform, but join on qualitatively equal terms as others.”⁷⁰ This means that the new entrant would not be “discriminated against by the dominant platform that might have its own competing service.”⁷¹ These authors refer to equitable interoperability as a “supertool” of platform regulation, providing many examples ranging from Internet Service Providers (ISPs) voluntarily joining the World Wide Web, to potentially opening up Google’s operating system and Apple’s app store.⁷²

The CERRE report also notes that interop can come in different degrees, at different layers, and in a continuum between full interoperability and no interoperability at all.⁷³ Interop can also be symmetric or asymmetric. Vertical interop can only be asymmetric — the platform gives one-way access to third parties — while horizontal can be either asymmetric or symmetric — the access goes in both directions.⁷⁴ Doctorow distinguishes between

68. *Id.* at 27.

69. *Id.* at 16.

70. Fiona M. Scott Morton et al., *Equitable Interoperability: The “Supertool” of Digital Platform Governance*, 40 YALE J. ON REGUL. 1013, 1016 (2023).

71. *Id.*

72. *Id.*

73. BOURREAU ET AL., *supra* note 63, at 16.

74. *Id.* at 16-17.

cooperative interop involving entities using the standard processes to voluntarily define and adopt interop, and adversarial interop involving entities plugging in their innovations without prior permission.⁷⁵ One can view the latter as resulting in a form of asymmetric interop.

For vertical interop, the CERRE authors suggest ways of defining and enforcing the access conditions. These range from giving the platform full autonomy to define the interfaces, to attaching certain non-discrimination terms to the access, to utilizing a multi-stakeholder process that defines the specifics of the interface, to formal standardization organizations that adopt common public standards. Platforms can also utilize measures such as access licenses to help ensure security, privacy, and other compliance requirements. A trusted third party or oversight board could develop the substance of the platform access rules.⁷⁶

C. Introducing Vertical AI Interop

As we have seen, interop can apply in a wide variety of contexts, including in and between different digital technologies. One form we have yet to see discussed widely is what could be called AI interop — connectivity that allows two or more advanced computational systems to communicate directly with one other. This form would take us above the more basic data and protocols layers to the entity's actual algorithm-based functions.

The Data Transfer Project's version of interop — allowing users to move their data between existing platforms — is one example of a positive related use case. However, it does not address what we would need from AI interop. We don't only seek permission to port our data from one large platform to another. We want to better enhance and promote our digital participation and agency. To accomplish this, interop is needed at the algorithmic layers, so the disparate islands and individual AI agents can connect to each other and/or to the underlying AI platforms.

Even ChatGPT seems to agree: "Lack of interoperability can hinder the deployment and integration of AI solutions, restrict collaboration between different stakeholders, and limit the potential benefits of AI. Interoperability ... fosters the development of a more connected and collaborative AI ecosystem, leading to

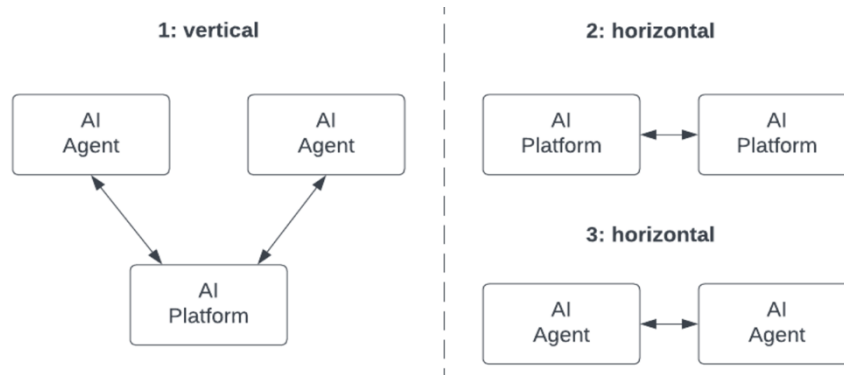
75. DOCTOROW, *supra* note 60, at 59-60.

76. BOURREAU ET AL., *supra* note 63, at 29-32.

improved efficiency, enhanced decision-making, and increased innovation.”⁷⁷

So where is AI interop in our marketplaces, our standards bodies, or our public policy conversations? Despite widespread acclaim for advances with generative AI, particularly LLMs, there is little public discussion about the notion of vertical interop between AI agents and Institutional AI systems. At best (and this is no small consideration), the right to individual contestability of certain decisions by AI systems has been introduced in Europe.⁷⁸

One can envision two primary modalities for interconnecting disparate AI networks. Vertical interop is connecting AI agents with larger AI platforms, while horizontal interop is connecting different AI platforms or AI-based agents.



Over the last twenty years, experts in the healthcare and IoT sectors have been honing a particular interoperability framework for the digital environment: the Levels of Conceptual Interoperability Model (LCIM).⁷⁹ According to one set of researchers,⁸⁰ the LCIM consists of six layers:

- **Technical:** the bit-passing layer of physical and virtual interconnection;

77. Interview with ChatGPT 4.0, OpenAI (Dec. 4, 2023) (responding to the query “Is AI interoperability valuable?”).

78. Margot E. Kaminski & Jennifer M. Urban, *The Right to Contest AI*, 121 COLUM. L. J. 1957, 1962 (2021).

79. See Dr. Andreas Tolk & James A. Muguira, *The Levels of Conceptual Interoperability Model*, in 1 2003 FALL SIMULATION INTEROPERABILITY WORKSHOP 53, (2003); Andreas Tolk, Saikou Y. Diallo, & Charles D. Turnitsa, *Applying the Levels of Conceptual Interoperability Model in Support of Integratability, Interoperability, and Composability for System-of-Systems Engineering*, 5 J. of SYSTEMICS, CYBERNETICS AND INFORMATICS 61, 66 (2007).

80. See Tolk et al., *supra* note 79.

- **Syntactical:** the data layer, defining the format of the data exchange;
- **Semantic:** the interpretation layer, adding understanding from common language and vocabulary;
- **Pragmatic:** the context layer, providing awareness of the other systems' methods and procedures;
- **Dynamic:** the state layer, capturing the effect of exchanged data on the Sender and Receiver; and
- **Conceptual:** the shared model layer, showing conformity with goals, purposes, and assumptions.

As can be seen, the LCIM provides new pathways for disparate systems to engage. In particular, the Pragmatic, Dynamic, and Conceptual Layers expand beyond the more linear data-centric focus of the lower OSI-inspired layers to include more rich and robust forms of machine-to-machine engagement.

To this basic framework, two additional layers can be added. First, one commenter suggests including an “Experiential Level” as the topmost layer.⁸¹ This would help confirm that complete interoperability benefits from users sharing both positive and negative experiences via a series of feedback loops.⁸² Second, as an early adopter of a form of AI interop, the healthcare industry has come to appreciate the human governance elements necessary to better move datasets and enable basic queries.⁸³ They have added what we can call the “Governance Level,” which serves as the human organizational layer to include social, legal, and economic considerations.⁸⁴ This suggestion also aligns well with the Interop authors, who devoted two of the four layers in their interop stack to the human and institutional layers.⁸⁵ “Very often, the most important parts needing to be rendered interoperable are humans and institutions, not technology and data.”⁸⁶ This overarching layer

81. Henrik Dibowski, *Semantic Interoperability Evaluation Model for Devices in Automation Systems*, in 22 IEEE INT'L CONF. ON EMERGING TECH. AND FACTORY AUTOMATION (2017).

82. *Id.*

83. See, e.g., John Kuhn, *AI And Large Language Models: The Future Of Healthcare Data Interoperability*, FORBES (June 20, 2023, 9:15 AM), <https://www.forbes.com/councils/forbestechcouncil/2023/06/20/ai-and-large-language-models-the-future-of-healthcare-data-interoperability> [https://perma.cc/R9RV-6MCN]; Oleksandr Telychko, *Semantic Interoperability in Healthcare*, CODEIT (June 1, 2023), <https://codeit.us/blog/semantic-interoperability-in-healthcare> [https://perma.cc/6WR5-342Y].

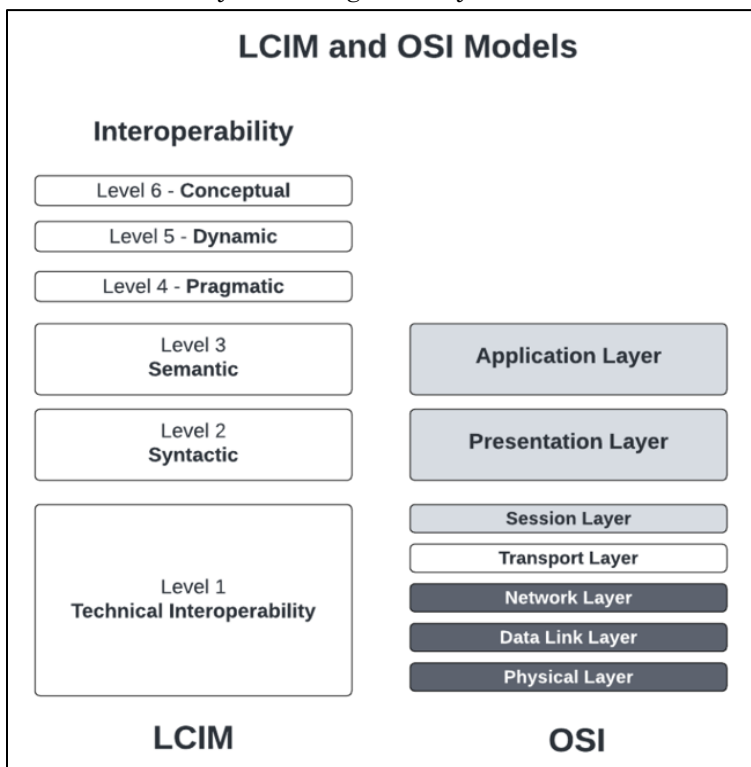
84. See generally, *supra* note 83.

85. PALFREY & GASSER, *supra* note 51 at 39-53.

86. *Id.* at 53.

also echoes the semi-facetious “Layers 8 and 9” of politics and economics added to the OSI stack by software engineering pioneer Evi Nemeth, as a way of highlighting the larger human context behind the protocols and software.⁸⁷

In the Web context, interoperability has allowed individuals and online entities to interact directly. For example, different kinds of Web browsers can access the same websites online. In the AI systems context, we need a similar function for each of us to actively engage with a variety of computational platforms. Adopting the modified interop stack outlined above could serve the core objective of robust agentic engagement.⁸⁸ What remains to be accomplished is devising the open standards and protocols necessary to enable vertical interop between digital agents and AI platforms.⁸⁹ Relatedly, devising ways for trustworthy intermediaries can provide the necessary robust agentiality.



⁸⁷. *Layer* 8, WIKIPEDIA, https://en.wikipedia.org/wiki/Layer_8 [https://perma.cc/FR2P-KMVT] (as of Nov. 17, 2024, 07:23 AM).

⁸⁸. Alice Gillin, *The role of AI in shaping the semantic layer: Insights from AtScale*, SILICONANGLE: *THECUBE* (May 23, 2023, 03:12 PM), <https://siliconangle.com/2023/05/24/role-ai-shaping-semantic-layer-insights-atscale-cubeconversations> [https://perma.cc/U5U9-6ASR].

⁸⁹. See Richard Whitt, *AI Interop: Roots, Benefits, Framework, Next Steps* (Jan. 2024), <https://www.glia.net> [https://perma.cc/D56C-SUYP].

D. Other Related Edgetech: SLMs and Middleware

The authentic Personal AI agent is but one example of technological innovations that can flourish at the “edge” of the Web. Unlike the “cloudtech” applications that are owned and controlled by Web platform companies and others, “edgetech” technologies can collectively be thought of as bringing more control to the end user. This tech typically involves storing personal data at the local level and using computational power at the device level. The proposed technology design standard that encompasses this way of thinking is the “e2a” (edge-to-any) standard.⁹⁰

By definition, all edgetech technology employs an interface that facilitates an edge-push and/or edge-pull functionality. Examples of such technology include edge computing, federated learning, OPAL algorithms, self-sovereign identity (SSI), personal data pods, and decentralized apps.⁹¹

Two other forms of edgetech that can be instrumental in the context of authentic PAI agents are Small Language Models and Middleware

1. Small Language Models

In the context of generative AI, another relevant innovation are language models that serve our personalized needs. LLMs are designed to train on enormous amounts of data drawn from the Web and elsewhere. By contrast, Small Language Models (SLMs)⁹² and Personal Language Models (PLMs) are built to represent individuals and small communities but can still connect and interact with LLMs. In the words of startup Personal.ai, “LLMs and PLMs can work seamlessly together to connect your internal memories with external knowledge.”⁹³

2. Middleware

A recent paper touts the creation of a new market of “middleware” providers that equip ordinary users with technologies

90. See generally WHITT, *supra* note 7, at 217-230.

91. Whitt, *supra* note 8, at 202-203.

92. David Razavi, *The Power of Small Language Models: A Quite Revolution*, MEDIUM (May 31, 2023), <https://medium.com/@davidrazavi/the-power-of-small-language-models-a-quite-revolution-baea7ac53d1b> [<https://perma.cc/HQ43-C2LK>].

93. *The Differences Between Personal Language Models and Large Language Models*, PERSONAL AI, <https://www.personal.ai/plm-personal-and-large-language-models> [<https://perma.cc/S44D-ANFC>] (last visited Nov. 17, 2024).

to manage their own interactions with social media networks.⁹⁴ Middleware refers to software products that can be appended to the major Web platforms. As its proponents explain:

These products would interconnect with Facebook, Amazon, Apple, Twitter, and Google APIs and allow consumers to shape their feeds and influence the algorithms that those dominant platforms currently employ. “Middleware would offer a third-party service chosen by the consumer and would make editorial judgments that are currently provided... by the platform.... Middleware can tailor the functionality of those websites to the preferences of their users.”⁹⁵

These authors presented the middleware concept in an amicus brief to the US Supreme Court⁹⁶ as a way to address the challenges posed by the application of Section 230 of the Communications Decency Act to large Web social media platforms.⁹⁷ The authors advocate for creating a new market for middleware, where individuals can substitute their own decision engines for the incumbent recommendation engines.⁹⁸ In essence, this middleware we can design for ourselves could swap out the decision engines provided to us on a “take it or leave it” basis by providing users with choices. Crucially, edgetech capabilities such as authentic PAI agents could be programmed to undertake the task of designing and implementing these individualized decision engines.

In all these instances, edgetech capabilities would provide an additional way for individuals and their authentic PAI agents to interact with other digital technologies in a more agential fashion. Taken together, these “e2a” technologies can form a robust ecosystem built on human agency. However, as with all technology, the overarching governance structures — who is actually in charge — loom large.

94. See Francis Fukuyama et al., *Middleware for Dominant Digital Platforms*, STANFORD: CYBER POLICY CENTER, https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/cpc-middleware_ff_v2.pdf [<https://perma.cc/47H7-QG9M>] (last visited Sep. 4, 2024).

95. *Id.* at 5.

96. Brief for Francis Fukuyama as Amici Curiae Supporting Respondents, *Moody v. NetChoice, LLC*, 143 S. Ct. 744 (2023) (No. 22-277) and Supporting Petitioners, *NetChoice, LLC v. Paxton*, 144 S. Ct. 275 (2023) (No. 22-555), https://www.supremecourt.gov/DocketPDF/22/22-277/292730/20231207155250803_22-277%20-555%20Fukuyama%20Amicus%20Brief%20Final.pdf [<https://perma.cc/QUP8-HAFU>].

97. See 47 U.S.C. § 230 (2018).

98. Fukuyama, *supra* note 94, at 9-14.

V. AGENTIALITY AND TRUSTWORTHY INTERMEDIATION

The crucial second dimension of agency is the authority to act on behalf of another. Today we have no uniform or widely acknowledged way to assess the agentiality of a purported individual-to-agent relationship on the Web. To date, the alignment approach seems to be limited to determining the scope of “human values” for purposes of training machine learning models.⁹⁹ While useful, this constrained focus is problematic because, as computer scientist Stuart Russell notes, it “seems inevitable ... that users will trust a personal assistant only if its primary obligation is to the user rather than to the corporation that produced it.”¹⁰⁰

A legal term I will be employing here is *privity*. To explain privity further, we will turn to a novel interpretation brought by Nick Szabo. In 1997, Szabo posted a prescient piece entitled “Recovering Privity,” which reconstitutes the legal concept of privity and applies it to the dawning information age.¹⁰¹

In recent years the term “privity” has been confined to contract law, in which it is defined as a relationship that exists between two contracting parties. This normally equates to the presence of a “meeting of the minds.” However, as Szabo notes, privity has far deeper roots than this “parched” modern day version.¹⁰² In the common law, the term privity was used to denote the creation of a boundary around two parties, defined by the scope of knowledge, consent, and control. Those within that boundary were said to be “in privity.”

The linguistic and semantic relation to “privacy” is no surprise. When adding control to privacy, privity becomes, in Szabo’s terms, a generalized “meta-relationship” where those within the bounds of the relationship are protected from those outside it.¹⁰³ For Szabo, the resulting “clarified boundaries of knowledge, control, and responsibility within and between relationships is ideal for specifying a variety of cyberspace relationships, whether informal

99. See, e.g., Oliver Klingefjord et al., *What are Human Values, and How do we Align AI to Them?*, MEANING ALIGNMENT INST. (Apr. 17, 2024) (posits utilizing Moral Graph Elicitation based on interviews with human participants about their values in particular contexts), <https://arxiv.org/pdf/2404.10636> [<https://perma.cc/6TGJ-N4PL>].

100. STUART RUSSELL, HUMAN COMPATIBLE: ARTIFICIAL INTELLIGENCE AND THE PROBLEM OF CONTROL 71 (2019).

101. Nick Szabo, *Recovering Privity*, UNIVERSITEIT VAN AMSTERDAM (1997), https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOT_winterschool2006/szabo.best.vwh.net/privity.html [<https://perma.cc/E3MF-86YT>].

102. *Id.*

103. *Id.*

or formalized via legal code or software.”¹⁰⁴ Here, we will refer to implementing and abiding by such a robust relational arrangement in the digital sphere as *Net privacy*.

Szabo’s resuscitated concept of privacy can be brought into the context of authentic PAI agents. Moving beyond the transactional mode of passive users and barely accessible Web entities, we can embrace the Net privacy of relationships founded on boundary layers of earned trust, mutual consent, and enhanced human agency. As the Google DeepMind paper helps illustrate, these boundary layers do not exist in a vacuum but rather intersect via relationships with those of other stakeholders, such as the developer/company and society at large.¹⁰⁵

As with agentivity, there are a number of potential indicators of agentiality in the online domain. For example, we can focus initially on the individual, who is ostensibly at the heart of the relationship. Various elements collectively can amount to a continuum establishing an authorized “meeting of minds” that exists between the individual and the developer/company, such that Net privacy actually exists. Indicia could include one or more of:

- Formal relationship, such as principal/agent;
- Explicit authorization, such as informed consent;
- Express exchange of mutual benefit, such as a contractual adequacy of consideration;
- Representational acts, such as signature, voiced approval, and other affirmative gestures;
- Clear communications of instructions, context, and tradeoffs;
- Structured independence from the legacy platform;
- Demonstrated compliance with the underlying terms of service or other unilateral instrument;
- Legitimate, objective basis for trust in the underling provider;
- Creation of autonomous spaces or designed frictions to accommodate an individual’s informed decision-making;
- Direct/indirect agentic benefits to the individual; and
- Adherence to any applicable legal requirements.

While none of these elements standing alone is a “silver bullet,” each deserves careful consideration when indicating whether a human is fully “on board” with decisions or actions being taken on their behalf.

104. *Id.*

105. Gabriel, *supra* note 43, at 41-43.

How can we best balance this individual-centric form of Net privity with the tetradic relationship proffered by the Google DeepMind paper? More specifically, how do we ensure the user's interests are not overridden with regulatory functions by the AI developer? As Bruce Schneier explains:

Imagine asking your chatbot to plan your next vacation. Did it choose a particular airline or hotel chain or restaurant because it was the best for you or because its maker got a kickback from the businesses? As with paid results in Google search, newsfeed ads on Facebook and paid placements on Amazon queries, these paid influences are likely to get more surreptitious over time.

If you're asking your chatbot for political information, are the results skewed by the politics of the corporation that owns the chatbot? Or the candidate who paid it the most money? Or even the views of the demographic of the people whose data was used in training the model? Is your AI agent secretly a double agent? Right now, there is no way to know.¹⁰⁶

Here is where the Google DeepMind concept of a tetradic relationship – balancing an AI's alignment between four potentially competing interests – can prove instructive. Rather than hiding behind the facile assumption that an AI company would always do what is in its users' best interests, the authors propose to explicitly surface the inevitable tensions that arise. These tensions exist whenever there may be a disparity between what an end user wants, and what the other stakeholders – the AI assistant, the developer, and society at large – might want. For the paper's authors, tetradic value alignment means finding that sweet spot between what may well amount to four competing claims to agentiality.

It is worth noting that the Google DeepMind paper assumes that "AI assistants will typically be aligned with the user's preferences or goals," as "most often," the corporate developers of the AI "aim to align the technology with the preferences, interests, and values of its [sic] users."¹⁰⁷ Nonetheless:

106. Schneier, *supra* note 25.

107. Gabriel, *supra* note 43, at 36-37.

[C]orporations have commercial objectives that exert independent force on the trajectory of a technology, states have national goals or priorities, and even independent developers may seek to further an ideological agenda or accrue reputational capital. These incentives may lead to the development of systems aimed at keeping users engaged or dependent ... or extracting information that can be used in other ways ... among other things.¹⁰⁸

The paper presents by way of example an all-too-likely scenario: the users want to adopt strong privacy protections that would limit AI developer access to “valuable” information. “This, in turn, might be commercially problematic insofar as it fails to generate a sustainable business practice,” so that “the AI systems might be value-aligned but not commercially viable.”¹⁰⁹ In this case, “there could still be a question about how to incentivize the development of this kind of technology to avoid socially costly ‘market failures’ and achieve real benefit.”¹¹⁰

It is a noteworthy advancement for a large Web/AI platform like Google to shine needed light on these perceived tensions between users and developers. But do these tensions actually exist? Are there other models where users and developers can remain well-aligned in a trustworthy relationship that also bring new market opportunities?

A. *A Pathway to Fiduciary AI Agents*

Some agree that the remedy for any agentiality gap is to ensure that the human being can be in charge of the relationship with the AI assistant and its developers. The Institute of Electrical and Electronics Engineers (IEEE), the leading global standards body, first launched its Global Initiative on Ethics of Autonomous and Intelligent Systems (A/IS) in 2016. That initiative’s inaugural paper highlights how digital agents rightfully should comport with what it calls “Ethically Aligned Design.”¹¹¹ IEEE has formally adopted eight governing principles for its initiative, including

108. *Id.* at 37.

109. *Id.*

110. *Id.* at 37 n.4.

111. INST. OF ELEC. AND ELEC’CS ENG’RS, Personal Data and Individual Agency, *in* ETHICALLY ALIGNED DESIGN (1st ed. 2019), https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e_personal_data.pdf [<https://perma.cc/679B-5PEM>].

human rights, well-being, data agency, and accountability.¹¹² Data agency signifies that people have “some form of sovereignty, agency, symmetry, or control regarding their identity and personal data,” while digital sovereignty is the ability “to own and fully control autonomous and intelligent technology.”¹¹³

IEEE goes on to endorse the A/IS agent as an end user’s educator, negotiator, and broker:

To retain agency in the algorithmic era, we must provide every individual with a personal data or algorithmic agent they curate to represent their terms and conditions in any real, digital, or virtual environment.... A significant part of retaining your agency in this way involves identifying trusted services that can essentially act on your behalf when making decisions about your data.... A person’s A/IS agent is a proactive algorithmic tool honoring their terms and conditions in the digital, virtual, and physical worlds.¹¹⁴

The Google DeepMind paper recognizes that human interpersonal relationships can provide useful analogies that carry over into the tetradic values alignment conversation.¹¹⁵ The paper acknowledges the existence of “power asymmetries” in a number of cases, such as the teacher-pupil, doctor-patient, and developer-user relationships. These asymmetries stem from the teacher/doctor/developer enjoying a superior position of authority and expertise over the pupil/patient/user.¹¹⁶ The paper further finds that “there seems to be a greater inherent power asymmetry in the human-AI case due to the *unidirectional* and *one-sided* nature of human-technology relationships.”¹¹⁷ More specifically, “power asymmetries can exist between developers of AI assistants and users that manifest through developers’ power to make decisions that affect users’ interests of choices with little risk of facing comparably adverse consequences.”¹¹⁸

What should be done to address these seemingly inherent power asymmetries between the companies developing AI

112. INST. OF ELEC. AND ELEC’CS ENG’RS, General Principles, in ETHICALLY ALIGNED DESIGN (1st ed. 2019), https://standards.ieee.org/wp-content/uploads/import/documents/other/ead1e_general_principles.pdf [https://perma.cc/N5TY-3GEV].

113. *Id.*

114. ETHICALLY ALIGNED DESIGN, *supra* note 111, at 114.

115. Gabriel, *supra* note 43, at 108–110.

116. *Id.*

117. *Id.* at 114 (emphasis in original).

118. *Id.* at 116.

assistants and those of us who will be using them? The Google DeepMind paper notes that “certain *duties* plausibly arise on the part of AI assistant developers,” duties that may be more extensive than those fiduciary duties that companies must abide by vis-à-vis their shareholders.¹¹⁹ Pointing again to cases where medical professionals and therapists who engage with “vulnerable individuals” are bound by fiduciary responsibilities, the authors intriguingly suggest that “a duty of care” might arise.¹²⁰ The paper hastens to add that even though “the moral considerations underpinning those professional norms plausibly apply to those who create these technologies as well, the same framework of responsibilities may not apply directly to the developers of AI assistants.”¹²¹ To which one can reply, “well, why not?”

Under the common law, fiduciary duties arose as a means of addressing power asymmetries between people.¹²² The doctrine is entwined with centuries of equity, torts, and other common law doctrine. Noted expert Tamar Frankel observed that “throughout the centuries, the problems these laws were designed to solve are eternal, etched in human nature, derived from human needs, and built into human activities.”¹²³

The basis for a fiduciary relationship is straightforward: assigning certain legal and moral obligations to people and entities engaged in exchanges of value with each other. The linchpin is what Frankel calls “entrusted power.”¹²⁴ An individual or entity (the entrustor, or beneficiary) grants access to something of value to another individual or entity (the fiduciary) for the purpose of having the fiduciary undertake tasks that benefit the entrustor. In these situations, the fiduciary normally has some knowledge, expertise, or other socially desirable capabilities that the entrustor lacks. Importantly, sensitive information is often revealed in the context of establishing the relationship (or even becomes its basis).

Not surprisingly, fiduciary law principles are near-universal, having been applied across a vast array of human endeavors. These include agency, trust law, corporate law, nonprofit law, pension law, employment law, bankruptcy, family law, international law, banking, health care, and public affairs. While most often associated with English common law, fiduciary law also

119. *Id.* (emphasis in original).

120. *Id.* at 116.

121. *Id.* at 116–117.

122. For an overview of these next several paragraphs, see generally Richard S. Whitt, *Old School Goes Online: Exploring Fiduciary Obligations of Loyalty and Care in the Digital Platforms Era*, 36 SANTA CLARA HIGH TECH. L.J. 75 (2020).

123. TAMAR FRANKEL, FIDUCIARY LAW 79 (2010).

124. *Id.* at 4.

encompasses most major global cultures—such as canon law, Roman law, classical Islamic law, classical Jewish law, European civil systems, Chinese law, Indian law, and Japanese law.

Prime modern-day examples of fiduciaries include the medical profession, the legal profession, and certain financial sectors. The entrustment of power to those providing these services triggers the obligation.

A number of scholars, such as Jack Balkin and Jonathan Zittrain, have proposed applying certain fiduciary duties to large Web companies as “information fiduciaries.”¹²⁵ Woodrow Hartzog and Neil Richards have proposed a similar notion of “data loyalty,” using fiduciary rights as a prism for defending strong privacy practices.¹²⁶ For the most part, these proposals have centered on protecting end users from harmful data practices and social media manipulations by the large Web providers. Most recently, Robin Berjon has suggested applying fiduciary duties to various user agents such as Web browsers, operating systems, and voice assistants.¹²⁷

The World Economic Forum (WEF), for one, has recognized several potential types of data intermediaries – including digital fiduciaries, data stewards, and data trusts – that can operate under fiduciary duties.¹²⁸ To date, however, few such entities exist in the world.

B. The GliaNet Initiative’s PEP Model

The GliaNet initiative, championed by the author, has introduced a new type of entity: a trusted digital intermediary. Trusted digital intermediaries provide technologies, like authentic PAI agents, that are answerable directly to us.¹²⁹ These “Net trustmediaries” (NetTMs) are companies or other entities that

125. See Jack M. Balkin, *The Fiduciary Model of Privacy*, 133 Harv. L. Rev. F. 11 (2020).

126. See Woodrow Hartzog & Neil Richards, *The Surprising Virtues of Data Loyalty*, 71 EMORY L.J. 985 (2022); Woodrow Hartzog & Neil Richards, *Legislating Data Loyalty*, 97 NOTRE DAME L. REV. REFLECTION 356 (2022).

127. ROBIN BERJON, *THE FIDUCIARY DUTIES OF USER AGENTS* (2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3827421 [<https://perma.cc/P8C2-55AD>].

128. WORLD ECONOMIC FORUM, *ADVANCING DIGITAL AGENCY: THE POWER OF DATA INTERMEDIARIES* (Danielle Carpenter, 2022), <https://www.weforum.org/publications/advancing-digital-agency-the-power-of-data-intermediaries> [<https://perma.cc/XGZ8-GXP9>].

129. See generally WHITT, *supra* note 7, at 165–213.

would be established to serve the individual as an actual client, and not a mere “end user.”¹³⁰

Importantly, these NetTMs act as “entrustors,” who voluntarily take on fiduciary law-based duties of care, loyalty, confidentiality, and good faith to each of its customers or clients who would be acting as willing “principals.” The GliaNet initiative has sketched out a plausible scenario – the “PEP Model” – for what an ecosystem of voluntary NetTMs would entail, using the common law of fiduciaries as a guide.¹³¹

First, the lowest level of obligation is a general duty of care. Entities could agree under law, or perhaps using a binding contractual agreement, to take reasonable and prudent steps to ensure that the PAI does not harm the individual. This could be seen as the *Guardian* role, to denote the role of protecting the customer. Examples in the digital environment include being transparent about your data policies, taking prudent steps to protect personal data from hackers, and refraining from selling or sharing one’s data with untrusted third parties.

Interestingly, the OpenAI white paper notes that allocating accountability for harms could include adopting the legal concept of a professional “standard of care.”¹³² In the Google DeepMind paper, this duty of care amounts to an obligation to “mitigate against user harms.”¹³³ While a necessary function, this seems to fall short of creating the more robust user alignment of values that the paper discusses.

Second, as scholars agree, the duty of care is not the core element of the fiduciary relationship: that obligation is the duty of loyalty. Loyalty can be defined in two different ways. The lower level – or “thin” duty of loyalty – entails no conflicts of interest or duties vis-à-vis other players.¹³⁴ This *Mediator* role can take on a *duty of fidelity*. With greater proposed agenticity in terms of what the PAI could accomplish on behalf of the individual, users could seek out entities who agree to take on such a duty. This would mean that the entity would mediate in an unbiased manner between the customer and various sources of news, information, commerce, and entertainment, perhaps by utilizing the middleware technology options discussed earlier.

Third, at the highest level, the individual could agree to be represented by a PAI that embraces “thick loyalty,” which requires

130. *See generally id.* at 75–184.

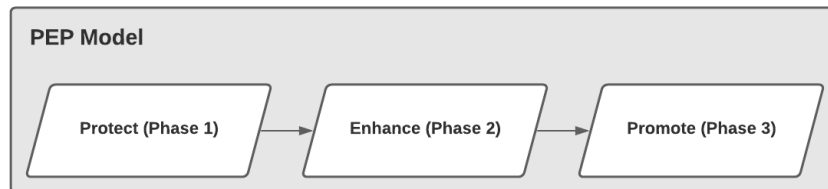
131. *See generally id.*, at 148–149.

132. SHAVIT, *supra* note 30, at 3.

133. Gabriel, *supra* note 43, at 117.

134. WHITT, *supra* note 7, at 135.

acting in the individual’s best interest. This *Advocate* role encompasses a straightforward *duty of loyalty*, where the entity actively promotes what is optimal for the user, and affirmatively informs them about online risks, filters out unwanted or harmful content from their online feeds, arms them with technology tools to protect themselves, and presents them with advertising and marketing options tailored to their particular wants or needs – or no ads at all.¹³⁵



One representation of the three phases of the Protect/Enhance/Promote model.

This “Protect/Enhance/Promote” (PEP) model is not just limited to variations on “user alignment.” Consistent with the tetradic approach suggested by Google DeepMind, the model can open up entirely new market opportunities, where individuals and entities can freely negotiate over the terms and conditions of digital representation.¹³⁶ The authentic PAI agent would be a crucial bridge between the user and the Web, taking on an increasing variety of agentic duties as the user’s duly authorized and trusted representative. As the capabilities of agenticity increase, so too would the depth of relationships of agentiality.

The Glianet initiative takes a two-tiered approach to Web governance.¹³⁷ Platform providers and others with access to user data would be regulated by law under the general duty of care (which corresponds to the “information fiduciaries” model championed by Jack Balkin).¹³⁸ By contrast, Net fiduciaries – a type of NetTM – would operate on a voluntary relational basis, pursuant to the higher duties of fidelity and loyalty towards their clients.¹³⁹ The premise is that mandating a general tort-like duty of care for those who interact with personal data is more politically and practically viable than mandating heightened duties of fidelity

135. *Id.* at 135.

136. *See id.* at 160-161.

137. *Id.* at 365-366.

138. *See generally* Balkin, *supra* note 125.

139. *See* WHITT, *supra* note 7, at 121-126.

and loyalty from those same entities. Or, as noted fiduciary expert Tamar Frankel puts it, forced loyalty amounts to no true loyalty at all.¹⁴⁰

The Google DeepMind paper professes to be “agnostic” on whether technology should only avoid reducing well-being or should actually improve it.¹⁴¹ That perspective is consistent with the Glinet initiative: the duty of care should apply as a matter of law in all instances, while market stakeholders can together explore technologies and business models that encompass the higher fiduciary duties of fidelity and loyalty.

Bruce Schneier has seconded the need to ensure AI is trustworthy by pairing it with trustworthy AI controllers, such as data fiduciaries. Applying this well-known societal system is “even more vital in a world of generative AI.”¹⁴² However, Schneier believes that markets of willing buyers and sellers on their own will not produce trustworthy digital agents and data fiduciaries.¹⁴³ That is a proposition well worth challenging before we resort to direct government regulation of all aspects of market behavior related to AI agents.

Importantly, the duties of fidelity and loyalty are not absolute. The common law of fiduciaries has always recognized and incorporated larger societal considerations. For example, physicians typically must identify highly contagious patients to health authorities, as well as report negligent acts observed by fellow physicians to licensing authorities. Additionally, under the “crime-fraud” exception, attorneys must report clients to law enforcement if they know there is a high probability their clients will commit a serious crime.¹⁴⁴ The real work will come with fashioning and applying an appropriate balancing test of interests between (at minimum) the four sets of stakeholders that Google DeepMind has identified – along with the trusted intermediary.

VI. ROBUST AGENCY: BETTER TOGETHER

Here is how the two dimensions of agency can play out functionally: vertical AI interops connect authentic PAI agents to

140. *Id.* at 249.

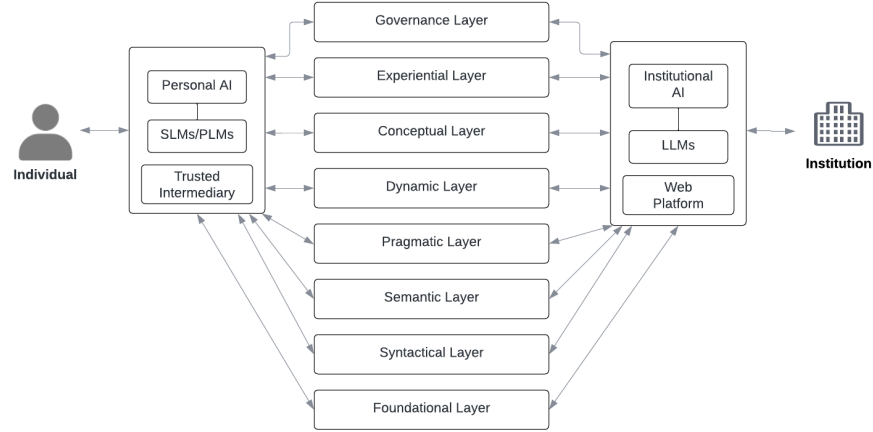
141. Gabriel, *supra* note 43, at 45.

142. Schneier, *supra* note 25; *see also* WILLIAM H. DAVIDOW & MICHAEL S. MALONE, *THE AUTONOMOUS REVOLUTION: RECLAIMING THE FUTURE WE’VE SOLD TO MACHINES* 129 (2020) (endorsing the “phase-change solution thinking” of information fiduciaries to protect individuals’ personal information).

143. Schneier, *supra* note 25.

144. WHITT, *supra* note 7, at 253.

other AI systems with NetTMs, such as Net fiduciaries, fully representing the empowered individual.



Put in terms of the edge-to-all (e2a) design principle mentioned earlier, vertical interop can serve two distinct purposes:

- *Edge-pull*: An individual can access AI platform resources as inputs, providing ways to choose external information and influences. This can be termed influence capture, or *freedom of impression*; and
- *Edge-push*: An individual can access AI-based platforms as a way to inject her human intentions into their decision engines (or substitute her own middleware). This can be termed intent-casting, or *freedom of expression*.

Both e2a functions enhance human agency and increase in value as more opportunities to promote intentions and capture influences become available. These opportunities would be realized on both the supply side (more decision engines are created and opened) and on the demand side (as those decisions become more personal and consequential).

Below are some potential use cases that demonstrate the robustness of this approach:

- *Protecting one's data flows*. The authentic PAI agent could manage the individual's privacy contours, the flow of personal data to online and offline environments. For example, this could include utilizing real-time machine-to-machine (M2M) connectivity to interpret a website's

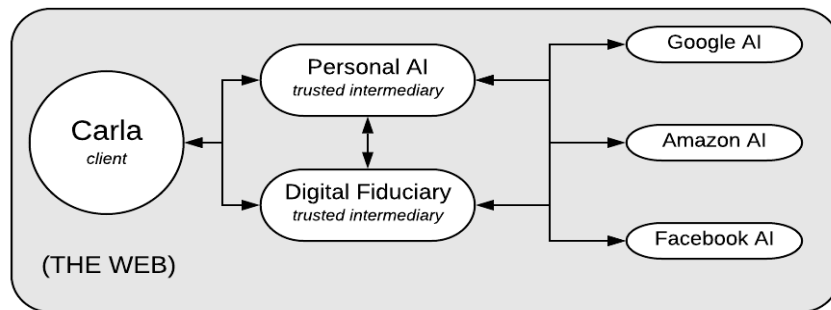
particular privacy policy and recommend whether the individual should or should not engage.

- *Broadcasting one's intentions.* The PAI could interpret a website or app terms of service (ToS) and other policies to generate tailored consent responses. One could even broadcast—"intentcast"—the client's own ToS to the Web, setting up interesting opportunities for real negotiation, bargaining, and compromise.
- *Creating and exporting one's own decision engines.* As noted above, the PAI could set individual preferences and defaults for the individual's dealings with news feeds, social networks, retail sites, and other online interfaces. The PAI also could ensure that Web-based recommendation engines are serving relevant information, and not harmful content, such as "fakes" or addictive videos. Those recommendation engines could even be replaced with "middleware" that better matches the individual's interests.
- *Circulating one's universal shopping cart.* The PAI could set up and manage one's own online shopping cart, pseudonymously taken from site to site. The cart could project only necessary identity information, using zero knowledge proof tech to validate its owner's financial viability.
- *Challenging the efficacy of consequential decision engines.* The PAI could contest financial, healthcare, law enforcement, and other impactful algorithmic systems for bias, false data, and other flaws that would harm the individual.
- *Mediating the terms of one's environmental engagement.* The PAI could help the individual dictate the terms of her immersive digital experiences with AR/VR/MR platforms. The PAI could even negotiate directly with environmental devices—smart speakers, facial recognition cameras, biometric sensors—and authorize, limit, or block access from surveilling and extracting personal data.
- *Building our own communities of interest.* The PAI could utilize these and other functions to enable the individual to construct her own actual and virtual communities, rather than those foisted upon her by others.

Importantly, in each case the primary focal point is the individual and their agential intentions. The human being should be considered the actual subject of their own digital destiny, armed with a full-fledged "PromoteMeBot." The individual human can be considered the center point of concentric rings that includes

relationships with their trusted intermediary, the original technology developer/company, the digital agent, and society at large. The trusted intermediary would help ensure that the individual is the actual subject of their own digital destiny. As the Google DeepMind paper shows, including all the relevant stakeholders is necessary for a comprehensive understanding of what we have been calling agentiality.¹⁴⁵ However, the proposal here is positing interwoven boundaries of influence that begin with the individual rather than a tetradic arrangement with no clear primary.

Here is one schematic for how an individual client, Carla, is represented by an authentic PAI agent and a trusted intermediary, which in turn utilizes agentic AI capabilities on her behalf with various Web/AI platforms.



Nonetheless, it is not clear that the digital agents currently being designed and released into the market in 2024 and beyond will meet those human agency-boosting aspirations. Consequently, there are a number of concerns.

First, the same Web companies who consider us mere “end users” are the primary suppliers of these agents. This means we technically are not even a customer, client, or patron to them. This lack of “Net privity” raises questions about the extent to which we have any exercisable rights over our relationship with these companies.

Second, one’s second-class citizenship as an “end user” extends to these companies’ terms of service (ToS). These documents, which are legally-binding on the individual end user, are unilaterally drafted and imposed via questionable consent processes. The ToS is also changeable at a moment’s notice without the need for prior consent. In terms of actual substance, at best the ToS grants the

145. See Gabriel, *supra* note 43, at 42-43.

user a revocable license to utilize the technology, secured by access to her personal data. However, these documents style the arrangement in which the user does not enjoy the full rights and privileges of ownership. Thus, the technology each of us is expected to utilize to represent our deepest selves is legally not even ours.

Third, the same companies that treat us as end users living under revocable licenses also tend to employ advertising-based business models premised on what I have called the “SEAMs” cycle paradigm (Surveillances, Extractions, Analyses, Manipulations). Under this paradigm, these entities *surveil* us as users, *extract* our personal data, *analyze* that data for insights into our thoughts and actions, and use those insights to *manipulate* our behaviors.¹⁴⁶ Clearly, the SEAMs cycles paradigm is flatly inconsistent with giving the individual more control over her online interactions.

Finally, as noted above, the most prominent digital agents are being provided to us by Google (Gemini), Microsoft (CoPilots), X (Grok), Amazon (Q), and OpenAI (GPTs). Most of these companies are fully integrated entities, from cloud to platform to AI model to data, and now to applications—or, in the case of OpenAI and Anthropic, have taken considerable financial investments from those same incumbents. As one set of commenters puts it, “[t]here is no AI without Big Tech.”¹⁴⁷ In their view, every startup, new entrant, and AI research lab depends on the big firms and their computing power, data sets, market reach, and other resources.¹⁴⁸ Even the promise of “open-source” AI models seems questionable, as unilateral licensing restrictions (again) can foreclose certain commercial opportunities. While vertical interop can enable competition, vertical integration threatens to limit our ability to move beyond their separate walled gardens to experience and benefit from the full range of AI-based capabilities.¹⁴⁹

We are then left with considerable doubts about our actual ownership of, and ultimate control over, these new digital agents. In reality, these agents look more like *pseudo*-personal AIs, or as Bruce Schneier memorably puts it, “double agents.”¹⁵⁰ Their true

146. WHITT, *supra* note 7, at 74.

147. Amba Kak et al., Opinion, *Make No Mistake – AI Is Owned by Big Tech*, MIT TECH. REV. (Dec. 5, 2023), <https://www.technologyreview.com/2023/12/05/1084393/make-no-mistake-ai-is-owned-by-big-tech> [<https://perma.cc/VJW8-6CX4>].

148. *Id.*

149. See Tejas Narechania & Ganesh Sitaraman, *An Antimonopoly Approach to Governing Artificial Intelligence*, VAND. POL’Y ACCELERATOR FOR POL. ECON. & REGUL. (Nov. 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4597080 [<https://perma.cc/2LFM-6E4V>].

150. Schneier, *supra* note 25.

owners would opt to use the full power of AI to understand us at the deepest levels, while providing us with services that are more responsive to themselves than to us.

The tetradic value alignment proposal found in the Google DeepMind paper is a useful starting point for assessing where and how the various fiduciary duties of care and loyalty should apply. Importantly, we should explore ways that the tetradic relationship can skew towards the developer at the expense of the user so we can try to address this problem.

VII. THE POLICY GAP

Normally, we would look to the world of public policy to understand the potential for useful change in laws and regulations. Fortunately, in 2023 and 2024 we have seen a range of public policy debates in response to the rise of generative AI. Unfortunately, those same debates have been incomplete.

A. *AI Accountability Agenda: Helpful, but Insufficient*

Against the backdrop of generative AI, governments around the world have sought “safe and responsible” algorithmic systems.¹⁵¹ Typically, this approach has included several components: a level of transparency, some “guardrails” to limit societal damage from high-risk use cases, and an enforcement regime. Without these algorithmic systems, lawmakers would be content to pave the way for the companies to bring their innovations to market. This approach is a linchpin to the EU AI Act,¹⁵² the Biden Administration’s Executive Order 14110,¹⁵³ the G7 principles,¹⁵⁴ the Bletchley Park Summit accord,¹⁵⁵ and US legislation.¹⁵⁶

151. See *infra* notes 149–152.

152. 2024 O.J. (L 1689).

153. Exec. Order No. 14100, 88 Fed. Reg. 75191 (Nov. 1, 2023).

154. *Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI system* (Oct. 30, 2023), <https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-guiding-principles-advanced-ai-system> [https://perma.cc/A3AE-KKFS].

155. Policy paper, *The Bletchley Declaration by Countries Attending the AI Safety Summit* (Nov. 1 2023), <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023> [https://perma.cc/JAY6-UWVP].

156. See Statement, Office of the High Commissioner for Human Rights, *Artificial intelligence must be grounded in human rights, says High Commissioner* (Jul. 12, 2023),

Binding or otherwise, these proposals can be instrumental in making Institutional AIs more responsible for their actions. However, they depend upon the same premise, partially modifying the ways that these AIs function while still leaving them largely under the incumbents' control. One can call this a behavioral approach to government regulation where laws seek to constrain what actions can and cannot be accepted.¹⁵⁷

While helpful in reducing the harmful behaviors that could be directed against us, this “AI Accountability Agenda” is not a complete formula unto itself. In particular, it fails to plant the public policy seeds necessary for developing alternate computational systems and business models that truly promote human agency and choice, multi-level market competition, and multi-layer tech innovations.¹⁵⁸ As we have seen, there is a complementary way to help facilitate these benefits by providing functional tools to allow new markets to form and potentially flourish.

B. Human Agency via AI Agenda

Rather than attempt to regulate the market behaviors of large AI incumbents, a Human Agency Agenda relies on devising and employing various ecosystem building tools. The Provisional Theory of Change (“PToC”) for the GliaNet initiative seeks to pave the way for additional players in the market to generate more competition, innovation, and user choice.¹⁵⁹ This approach combines the two elements outlined above: human empowering governance with human empowering “edgetech” applications and services. As this paper has outlined, this entails, among other things, facilitating the two dimensions of agenticity via vertical AI interop and agentiality via authentic relationship.

Arguably, the U.S. Congress has already supported the spread of edgetech-style tools, such as authentic PAI agents. In an often overlooked subsection of Section 230 of the Communications Decency Act, Congress includes a finding that interactive computer services operating via the Internet “offer users a great degree of control over the information that they receive, as well as the potential for even greater control in the future as technology

<https://www.ohchr.org/en/statements/2023/07/artificial-intelligence-must-be-grounded-human-rights-says-high-commissioner> [<https://perma.cc/9KJS-8SH7>] (even the UN High Commissioner for Human Rights, in critiquing the EU AI Act, focused on reducing harms, rather than empowering human rights in the digital world).

157. See Whitt, *supra* note 49, at 76.

158. *Id.*

159. WHITT, REWEAVING THE WEB, AT 231–34.

develops.”¹⁶⁰ The statute builds on this finding to declare that “it is the policy of the United States ... to encourage the development of technologies which maximize user control over what information is received by individuals, families, and schools who use the Internet. . . .”¹⁶¹

Clearly, many “edge-pull” tech tools, like the ones described above, meet that policy objective. In fact, pursuant to a recent lawsuit against Meta by Ethan Zuckerman, a federal court has been asked to recognize that Section 230 protects the development of software tools that empower social media users to control what they see online.¹⁶² An authentic PAI agent empowered by robust vertical interop can become the trusted mediator for maximizing such user control over their online interactions with other AI systems, including social media platforms.

Consistent with Section 230, the key levers to an AI agency agenda would be public policies that encourage both the technology and governance dimensions of AI systems. As we have previously touched on, an open and secure interop would seek to preserve the benefits of both while also minimizing the downside. But how exactly would this come about?

VIII. SOME NEXT STEPS

Recognition of this gap between the external packaging and the reality of “personal” AI “agents” point to the need for authentic digital agency to be accountable to the individual. Given the vertical integration already taking place, where can we hope to find these wholly independent and human-centric digital agents?

Pursuing both the goals of achieving AI interop, for the agentic dimension, and trustworthy intermediation, for the agential dimension, will ensure that each is given appropriate weight in the overall AI governance landscape. With vertical AI interop, open standards and protocols can be pursued through multistakeholder bodies, such as the IEEE and IETF. With trustworthy intermediaries, interested entities can band together to create professional codes of conduct and other explicit indicia of adherence to fiduciary law-based duties.

160. 47 U.S.C. § 230(a)(2).

161. 47 U.S.C. § 230(b)(3).

162. Zuckerman v. Meta Platforms, No. 3:24-cv-02596 (N.D. Cal.); *see also* Zuckerman v. Meta Platforms, Inc., KNIGHT FIRST AMEND. INST. AT COLUMBIA UNIV. (Aug. 29, 2024), <https://knightcolumbia.org/cases/zuckerman-v-meta-platforms-inc> [<https://perma.cc/TP87-JTRR>].

With our objective in mind—using vertical AI interop as a means of promulgating authentic digital agents and trustworthy intermediaries to promote human agency—how can we take concrete steps to move forward? Three avenues are worth touching on: software standards, corporate policies, and public policy.

A. *The Role of Software Standards*

1. Open Interop APIs: A Sufficient Starting Point?

An initial approach to approximating vertical AI interop could be the use of open Application Programming Interfaces (APIs). These software interfaces have been utilized with much success across a wide swath of the digital economy.¹⁶³ The Interop authors laud open APIs as being potentially powerful drivers of innovation.¹⁶⁴

In the spaces where these APIs can approach the functionality of interoperability, they should at least be useful stop-gap options. However, in the context of AI systems, that similarity in functionality is unclear.

APIs exist for the purpose of accessing a platform's resources to build something new. They do not exist for the purpose of conveying a client's intentions to gain specified actions. They are more a means of pulling in resources rather than pushing out intentionalities. A typical API, if employed to access an AI system, would allow a third party to build a digital agent. But, in the absence of another form of interop, it would be tied inextricably to that platform. Generally, each platform can build an API to its own specifications.

“Web APIs” live on the classic HTTP standard and client-server model of cloud tech as developer interfaces to a platform's websites and applications. This asymmetric arrangement lets the API owner dictate the terms to those who seek to utilize it, in which those terms can change unilaterally. Would “AI APIs” look and act any differently?

An important consideration is the unilateral nature of these interfaces, rendered by entities with evolving motivations. As New America points out, vertically integrated platforms have incentives to build their API design solely to their own needs, tailored to their own specific apps, features, and competitive strategy.¹⁶⁵ Indeed,

163. See *generally* OPENAPI, <https://www.openapis.org/about> [https://perma.cc/8EQE-KLG6] (last visited Sept. 7, 2024).

164. PALFREY & GASSER, *supra* note 51, at 116.

165. Chao & Schulman, *supra* note 50, at 15–16.

By setting API design and policy, [online platforms] have the ability to control who has access to critical aspects of the vast datasets and user bases they've built—things like a user's social graph that enables a hopeful competitor to grow its own user base and establish itself. Once a platform is sufficiently scaled, and especially if it is dominant, it no longer has the incentives to grant access to its APIs to facilitate a healthy downstream ecosystem. The more vertically integrated a platform is, too, the higher the risk that it may not offer APIs with sufficient data and functionality for other companies. Whereas our current antitrust framework may not sufficiently ensure platform competition, platform interoperability offers a solution to promote a more competitive ecosystem.¹⁶⁶

By contrast, Element believes, “[a]n open standard means that anyone and everyone can invest in development—driving innovation, raising standards, creating jobs—safe in the knowledge that the ground won't move from under them. Knowing that Big Tech can't simply whip away the API rug.”¹⁶⁷

In sum, open APIs may be a stepping-stone to broaden access to API systems, but vertical interops are needed to provide the necessary tools for a robust market of independent digital agents.

2. Open Standards For Digital Agents

What is necessary to bring alive this vision of a robust ecosystem of authentic digital agents is the software code linking together the underlying AI systems. To date, there are no open standards or protocols to support vertical AI interop. Thus, the term “open” can be fraught in technology circles.¹⁶⁸

As noted above, IEEE's Global Initiative on Ethics of Autonomous and Intelligent Systems has created some important intellectual groundwork for authentic digital agents that fully represent the autonomous human being.¹⁶⁹ At the IETF, a new initiative is underway that may be picking up where the IEEE's

166. *Id.* at 12.

167. Amandine Le Pape, *Interoperability: Open APIs Are a Start, Open Standard is Better*, ELEMENT (Feb. 28, 2022), <https://element.io/blog/interoperability-open-apis-are-a-start-open-standard-is-better-2> [<https://perma.cc/K4D6-JNQJ>].

168. *See generally* Whitt, *supra* note 49.

169. *See generally* INST. OF ELEC. & ELEC'CS ENG'RS, ETHICALLY ALIGNED DESIGN (1st ed. 2019), https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e_personal_data.pdf [<https://perma.cc/4J67-J4JJ>].

P7006 working group has left off. The “Personal Digital Assistant Protocol” (PDAP) is based on the Universal Human Right of Freedom of Association and Assembly, which translates to the individual choice of hosting and supporting communities for one’s digital agent.¹⁷⁰ Its overarching purpose is to enable a shift from proprietary platforms to personal agents. These agents would be interoperable by vendors and service providers, avoid lock-in to hosts, give individuals a choice of community to replace forced platform association, and allow the agent’s policies to be portable across host communities.¹⁷¹ A key technical component is allowing a separate choice of the Authorization Server as agent from the Resource Server as vendor.¹⁷² The protocol would also standardize Request Presentation and Authorization Tokens, as well as the Service Endpoint for a personal digital agent Authorization Server.¹⁷³

Even if the IETF-sponsored work is successful and related efforts are launched, more networking and interfaces standards development will be necessary to move us to a world of robust AI interoperability.

B. Relying on Market Adoption

1. Voluntary Interop?

Entities seeking to serve their clients as trustworthy intermediaries can develop software standards and protocols to facilitate vertical AI interop. Obviously, these practices will mean little if they are not adopted and successfully implemented in the marketplace.

Some Web platforms may believe that voluntarily adopting vertical AI interop could open vast new markets while also forestalling calls for more direct government regulation of their business practices. As one set of commenters notes:

First and foremost, it is part of the entrepreneurial freedom of firms to decide themselves on the extent of the interoperability of their products and services. Selling products that are interoperable with other products, or offering an open platform that allows for sharing products

170. Internet Engineering Task Force, *Personal Digital Agent Protocol* (Nov. 5, 2023), <https://datatracker.ietf.org/meeting/118/materials/slides-118-hotrfc-sessa-10-personal-digital-agent-protocol-01> [<https://perma.cc/7YX3-LZAU>].

171. *Id.*

172. *Id.*

173. *Id.*

and services with other platforms, can increase the value for customers and therefore increase profits. In the same way, the use of standardized components in a production value chain can reduce production costs and therefore allow for lower prices.¹⁷⁴

For example, the Data Transfer Initiative could be an industry forum for leading such activities.¹⁷⁵

A group of commenters have contributed an important concept worth fleshing out in the context of pro-interop corporate policies. A Design for Interoperability, or “DfIOp,” can build in digital interoperability as a core component of future product design.¹⁷⁶ This would include an outer interoperability between the system’s boundaries and its environment of third-party agents and applications.

2. Voluntary Association of the Willing?

As noted above, the proposed GliaNet ecosystem combines: (1) imposing a general duty of care on entities that have access to and use our personal data and (2) facilitating a new market for entities agreeing to adopt duties of fidelity and loyalty towards their clients. The latter element depends on for-profit companies and other entities seeing the value in establishing fiduciary relationships that encompass those duties by instantiating them in their product and service offerings. The Theory of Change here is that PEP-informed relationships built on genuine trust and protection will generate at least as many—if not more—financial benefits than the current Web ecosystem built on the SEAM cycles of surveillance, extraction, analysis, and manipulation.¹⁷⁷

The single best way to prove the efficacy of this Net™ mode is to create it. As of September 2024, the GLIA Foundation has launched the GliaNet Alliance, a coalition of companies, investors, advisors, and others dedicated to building a new marketplace of

174. Wolfgang Kerber & Heike Schweitzer, *Interoperability in the Digital Economy*, J. INTELL. PROP. INFO. TECH. & ELEC. COM. L. 39, 42 (2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2922515 [<https://perma.cc/J7TL-6TKM>].

175. Interestingly, the Data Transfer Initiative—picking up from the now-defunct DTP—recently suggested that interop could be a key to unlocking the value of LLMs. See Chris Riley, *The Future of Generative AI is Personal – and Portable?*, DATA TRANSFER INITIATIVE (Nov. 21, 2023), <https://dtinit.org/blog/2023/11/21/future-AI-portable> [<https://perma.cc/EK6C-DEQB>].

176. See MARTIN SJAROV ET AL., *Towards “Design for Interoperability” in the Context of Systems Engineering*, 96 PROCEDIA CIRP 145, 147–48 (2020).

177. WHITT, REWEAVING THE WEB, AT 343–53.

trustworthy digital relationships.¹⁷⁸ Time will tell whether such a coalition, and others alike, can move the proverbial needle sufficiently to represent a true market-oriented alternative.

Others see the opportunity in bringing fiduciary values to those who build AIs. For example, Chinmayi Sharma has introduced a project to issue professional licenses to AI engineers. As she puts it, “What if, like doctors, AI engineers also vowed to do no harm?”¹⁷⁹ Whatever challenges such a proposal faces, there is little doubt that adopting a customary standard of care for the software engineers building AI technologies “would ensure that the only people building AI are those who are both qualified to do so and are doing so in sanctioned ways.”¹⁸⁰ One intriguing aspect of Sharma’s proposal is the use of a customary standard of care as a legal shield against malpractice claims for the negligent production of harmful software.¹⁸¹ This would be an example of creating an attractive incentive for individuals and entities to agree to become “professionalized” AI engineers.

C. Public Policy: Codifying Our Rights and Their Duties

1. The Current Landscape for Interop

As we have seen recently with OpenAI, relying solely on self-governance can be a shaky basis for guaranteeing one’s digital rights. In the case of vertical AI interop, some Web platforms may resist adopting the necessary standards and protocols. As a result, we may need to consider forms of government intervention to make interop a reality. Importantly, as we have seen:

[I]nteroperability is not—or should not be—an end in itself; it is a means to a broader set of goals: to address market fragmentation, to avoid market tipping towards monopoly; to open downstream markets for competition where the upstream market is monopolized; to increase follow-on innovation irrespective of market power; or to address a perceived societal need for general interconnectedness and communication across competing networks. In each case,

178. *Building a More Trustworthy Web*, GLIANET, <http://www.glianetalliance.org> [<https://perma.cc/ZX59-37GR>] (last visited Nov. 3, 2024).

179. Chinmayi Sharma, *AI’s Hippocratic Oath*, WASH. U. L. REV. (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4759742 [<https://perma.cc/64V7-MW2F>].

180. *Id.* (manuscript at 35).

181. *Id.* (manuscript at 35–43).

before taking action, a clear and strong market failure or public service rationale should be identified.¹⁸²

In other words, the government's direct involvement requires sufficient evidence that interoperability as a public good will not occur otherwise.

In Europe, interoperability has become a key component of the legal landscape.¹⁸³ Competition law has incorporated interop as a remedy to address purported abuses of dominance under Article 102 of the Treaty on the Functioning of the European Union (TFEU).¹⁸⁴ More recently, the EU's Digital Markets Act (DMA) of 2020 introduced new obligations for core platform providers or "gatekeepers." Article six makes clear that interoperability is one way to address concerns about various self-preferencing behaviors by attaching operating systems with third party software applications, stores, and to ancillary services. Article six also specifies real-time data portability and gives business users the right to access their own end user data.¹⁸⁵

Nonetheless, no explicit interop right exists today for AI systems. Given this gap, and the proposed benefits of such a right, one option is for policymakers to step in. One approach is to utilize various "soft power" inducements to stimulate the adoption of vertical interop to support the growth of new markets. These inducements could include procurement preferences, tax breaks, and government grants.

Chris Riley, then with Mozilla Corporation, proposed an intriguing option. As noted above, Section 230 of the U.S. Communications Decency Act provides immunity to online companies that host user-generated content. In recent years, that provision has come under political attack by those who believe the larger social media companies and other online entities have done an insufficient job of moderating that content. Riley's proposal conditions continuing immunity under Section 230 for those large

182. Kerber & Schweitzer, *supra* note 174, at 58.

183. BOURREAU ET AL., *supra* note 63, at 36.

184. Consolidated Version of the Treaty on the Functioning European Union art. 102, Oct. 26, 2012, 2012 O.J. (C 326) 47, 89.

185. Proposal for a Regulation of the European Parliament and of the Council on Contestable and Fair Markets in the Digital Sector (Digital Markets Act) art. 6 (1)(c); (1)(f); (1) (h); 1(i), COM (2020) 842 final (Dec. 15, 2020); *see also* BOURREAU ET AL., *supra* note 63, at 37–39.

online intermediaries who provide “an open, raw feed for independent downstream presentation.”¹⁸⁶

Riley’s thesis is that centralizing and siloing one’s Internet experience has produced “a lack of meaningful user agency.” “There are fundamental limits on users’ ability to customize, or replace, the recommendation algorithm that mediates the lion’s share of their interaction with the social networks and the user-generated content that it hosts.”¹⁸⁷ Interoperability would allow end users to customize their online experiences and experiment with substituting new types of presentation algorithms or other “middleware” options. Thus, Riley’s proposal would allow platforms to adopt reasonable security and privacy access controls for their APIs or other interop interfaces.¹⁸⁸

2. The Current Landscape for Agential Relationships

As much as vertical AI interop is essential to enhance agentic capabilities, creating agential relationships requires empowering authentic PAI agents to act on our behalf. As we have seen, one way to accomplish this is to confer users a right to contest AI decisions.¹⁸⁹ As its proponents note, the government and private sector use algorithms to “decide how to distribute welfare benefits, whether to hire or fire a person, whether expressive material should be removed from online platforms, whether to keep people in prison, and more.”¹⁹⁰ One can think of these as consequential decisions as being rendered in whole or in part by algorithmic systems. The ability to contest an adverse decision rendered by an AI system should be an important right.

In the context of personal data, such a right does exist in Europe. The GDPR applies to the processing of personal data by both the government and the private sector. Article twenty-two of the GDPR states that for certain automated decisions, affected individuals must be provided “at least the right to obtain human intervention . . . to express his or her point of view and *to contest the decision*.”¹⁹¹

186. Chris Riley, *Section 230 and Interoperability*, MEDIUM (Sept. 10, 2020), <https://mchrisriley.medium.com/section-230-and-interoperability-2d63e225088d> [<https://perma.cc/59GK-NW8Y>].

187. *Id.*

188. *Id.*

189. *See generally* Kaminski & Urban, *supra* note 78.

190. *Id.* at 1960.

191. Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data art. 22, 2016 O.J. (L 119) 1, 46.

Nonetheless, correcting another's assumedly flawed decision-making does not by itself imbue the human with optimal agency. The system is still making decisions on our behalf. Substituting one's own decision, as with the middleware proposal outlined above, would take us one step closer to human agency. Perhaps most transformational would be the ability to actively promote one's own decision engines in the first place. These options will only be possible when authentic digital agents are able to interoperate directly with the AI platform in question.

If necessary, in the face of demonstrable market failure, policymakers also can move into "hard power" mode by fashioning laws and regulations with express requirements. As just one example, in the United States, the ACCESS Act, if adopted, would help turn these proposed standards and codes of conduct into actual legal requirements.¹⁹² This bill combines recognition of "custodial" third parties operating on behalf of users under specified care-like duties with provisions mandating forms of data portability and interoperability. Senator Mark Warner, the bill's lead sponsor, explained that:

"[E]mpowering trusted custodial companies to step in on behalf of users to better manage their accounts across different platforms will help balance the playing field between consumers and companies. In other words—by enabling portability, interoperability, and delegatability, this bill will help put consumers in the driver's seat...."¹⁹³

Noted economist Paul Romer agrees. In his own supporting statement, Romer observed that "by giving consumers the ability to delegate decisions to organizations working on their behalf, the ACCESS Act gives consumers some hope that they can understand what they are giving up and getting in the opaque world that the tech firms have created."¹⁹⁴

IX. A PROPOSED RIGHTS/DUTIES FRAMEWORK

Whether adopted through the governance options of open standards, corporate policies, public policy inducements, and/or mandates, much remains to be decided to create a vertical AI interop and trusted intermediaries framework. As noted in the

192. See Press Release, Mark R. Warner, U.S. Senator for Va., Senators Introduce Bipartisan Bill to Encourage Competition in Social Media (Oct. 22, 2019), <https://www.warner.senate.gov/public/index.cfm/2019/10/senators-introduce-bipartisan-bill-to-encourage-competition-in-social-media> [<https://perma.cc/A4XJ-C7QF>].

193. *Id.*

194. *Id.*

Google DeepMind paper, many open questions remain, including exactly who gets what rights and duties, under what circumstances, and for what purposes.

Below is one proposed “straw person,” suitable for any of these governance options. This proposal could be called a *Human:Digital Code of Rights and Duties*. One can think of this proposal as encompassing the four elements of “MIDI”—the right to *mediate*, based on the functions of *interoperability*, *delegation*, and *interrogation*. Each of those four elements applies to an individual person—the “me.” The GLIA Foundation submitted a similar proposal to the European Commission in June 2020,¹⁹⁵ in response to the EC’s February 2020 white paper on AI.¹⁹⁶

Our Human:Digital Rights

- *Delegation*: The ability to delegate my *human:digital* rights to any designated third-party mediating entities.
- *Clarity*: The ability to understand the impact of third-party computational systems on me.
- *Query*: The ability to ask relevant questions and receive relevant responses about consequential decisions.
- *Negotiate*: The ability to engage in give-and-take over the terms and conditions of such decisions and impacts on me.
- *Challenge*: The ability to register concerns and opposition to consequential decisions and impacts on me.
- *Recourse*: The ability to seek and gain recourse for the impact of such decisions on me.
- *Consent*: The ability to provide meaningful, voluntary, express consent to decisions of consequence.
- *Substitute*: The ability to supply one’s own decision engine in place of that preferred by the third-party computational system.

“Decisions of consequence” could be defined as any human or machine decision with a potentially consequential impact on a person’s life, including health care, employment, legal status, financial capability, citizenship, property or contract rights, ownership of assets, freedom of movement, and access to government benefits.

195. Memorandum from the GLIA Found. to the European Comm’n, How to Create an ‘Ecosystem of Trust’: GLIA Foundation’s Proposals on the European Commission’s White Paper on Artificial Intelligence (June 13, 2020).

196. See EUROPEAN COMM’N, WHITE PAPER ON ARTIFICIAL INTELLIGENCE (Feb. 19, 2020).

Their Duties

These requirements would apply to any entities employing computational systems with data, algorithms, and interfaces that significantly affect my ability to exercise my human digital rights and render decisions of consequence about myself.

Duties of institutional accountability

- *Overarching duty of care*: do not harm me.
- *Overarching duty of good faith*: do not treat me as a mere object.
- *Overarching duty of confidentiality*: do not harm my personal contours of privacy.

Duties of human agency

- *Interoperate*: connect your physical and virtual systems with my designated entities.
- *Portability*: move my data to me/my designated entities.
- *Transparency*: tell me exactly what you are doing to me.
- *Response*: answer my queries.
- *Deference*: honor my requests and decisions.
- *Provide optionality and recourse*: give me relevant options.

CONCLUSION

As a follow-on to his 1988 Knowbots paper with Bob Kahn, Vint Cerf six years later penned for IETF a whimsical RFC 1607, entitled “A Visit from the 21st Century.”¹⁹⁷ In it, Cerf imagines a future world where the KnowBots have become prevalent. Among other advances, their control subsystems now allow the “NanoSystem” to be fully linked to the Internet. This allows the patron to control data sharing and other forms of inter-system communication.¹⁹⁸

As our Web interfaces continue to recede away from our direct control—from online screens, to embedded IoT scenes, and now to the “unseens” of advanced cloud-based algorithmic systems—technology advances like generative AI platforms can enhance our digital selves in myriad ways. The key is who is actually in charge. As RFC 1607 suggests, fully connected and user-controlled AI systems can engender a welcome future of authentic digital agents—essentially, the PromoteMeBots. Similarly, under the GliaNet PToC, a combination of vertical interoperability and

197. See generally Vinton G. Cerf, *A Visit from the 21st Century*, INTERNET SOCIETY (Apr. 1, 1994), <https://www.rfc-editor.org/rfc/rfc1607.html> [<https://perma.cc/3353-5HTD>].

198. *Id.* at 3.

trusted intermediation holds open the potential to shift empowerment from the cloud to the edge. That said, as the Interop authors note, no single application or configuration of actors can create an optimal level of interop in every case. “The converse is also true: there are many ways to get to interoperability, many actors who can participate, and many tools in the toolbox that the full range of actors can put to work.”¹⁹⁹

With widespread adoption of vertical AI interop and trusted intermediation, each of us can enjoy innovations such as authentic Personal AI agents and other edgetech employed for us by Net fiduciaries. The full futuristic vision of the PromoteMeBot can then be realized, with ordinary people exploring ways of flourishing—on their own terms.

199. PALFREY & GASSER, *supra* note 51, at 160.