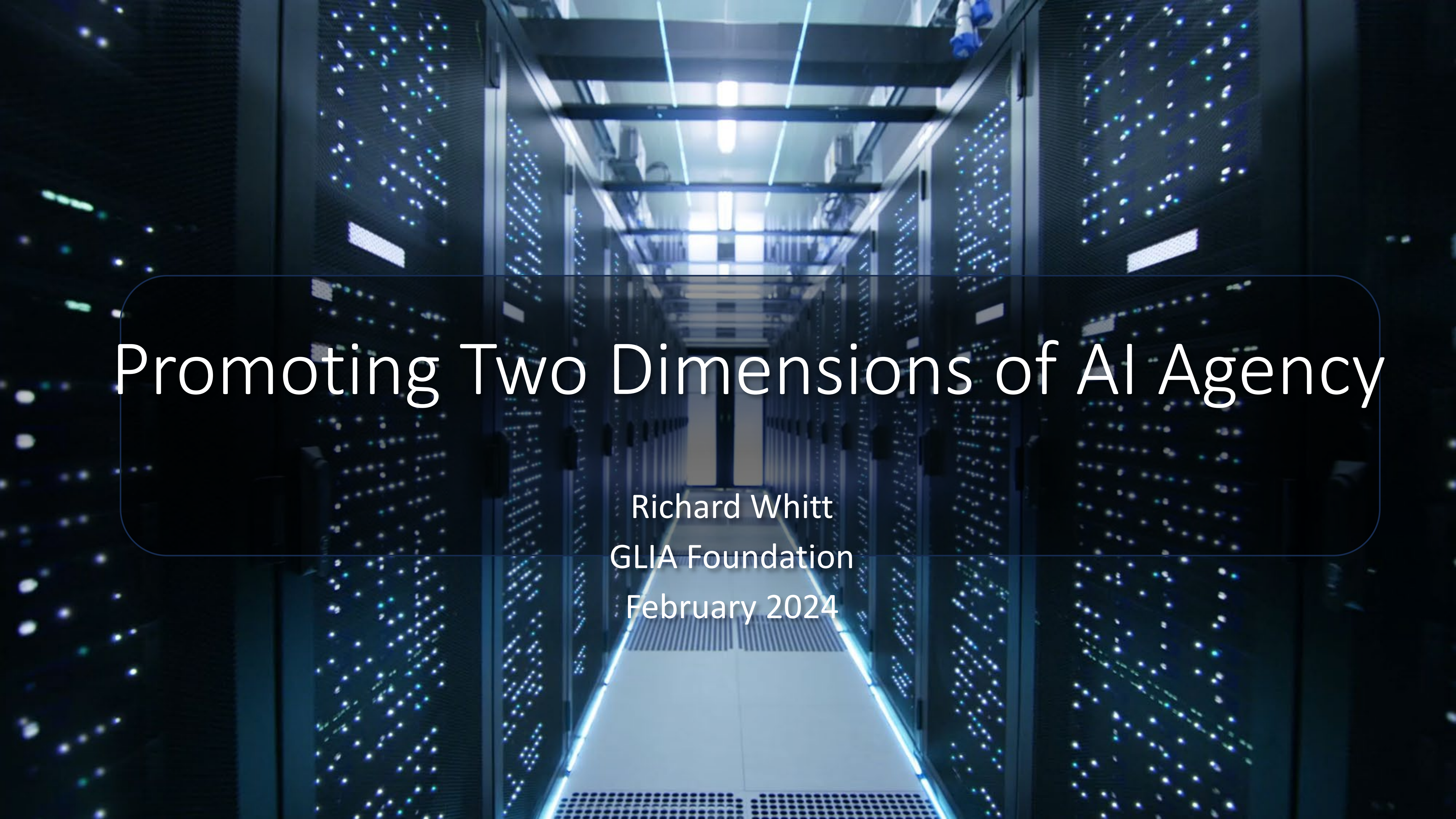
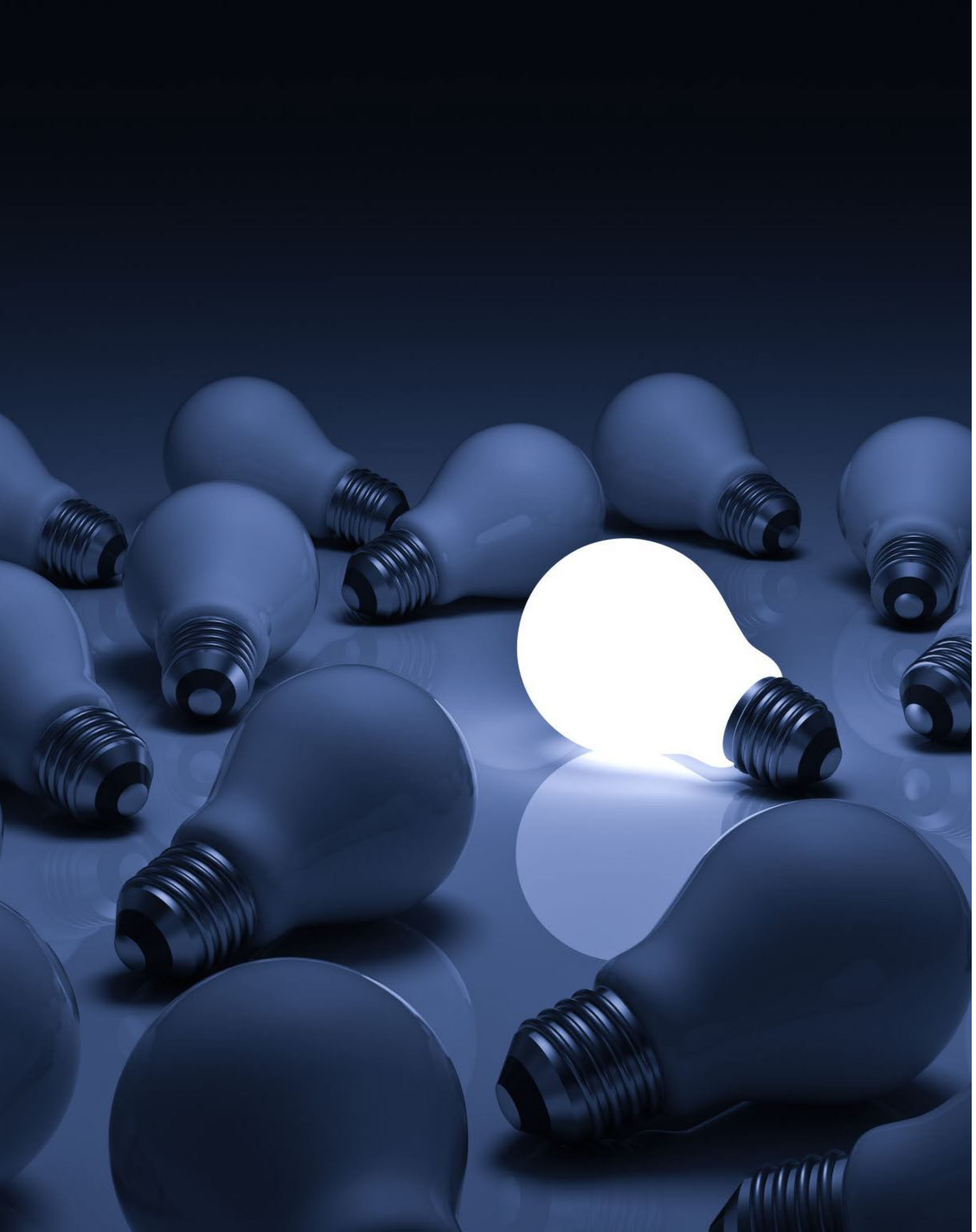


Promoting Two Dimensions of AI Agency

Richard Whitt
GLIA Foundation
February 2024



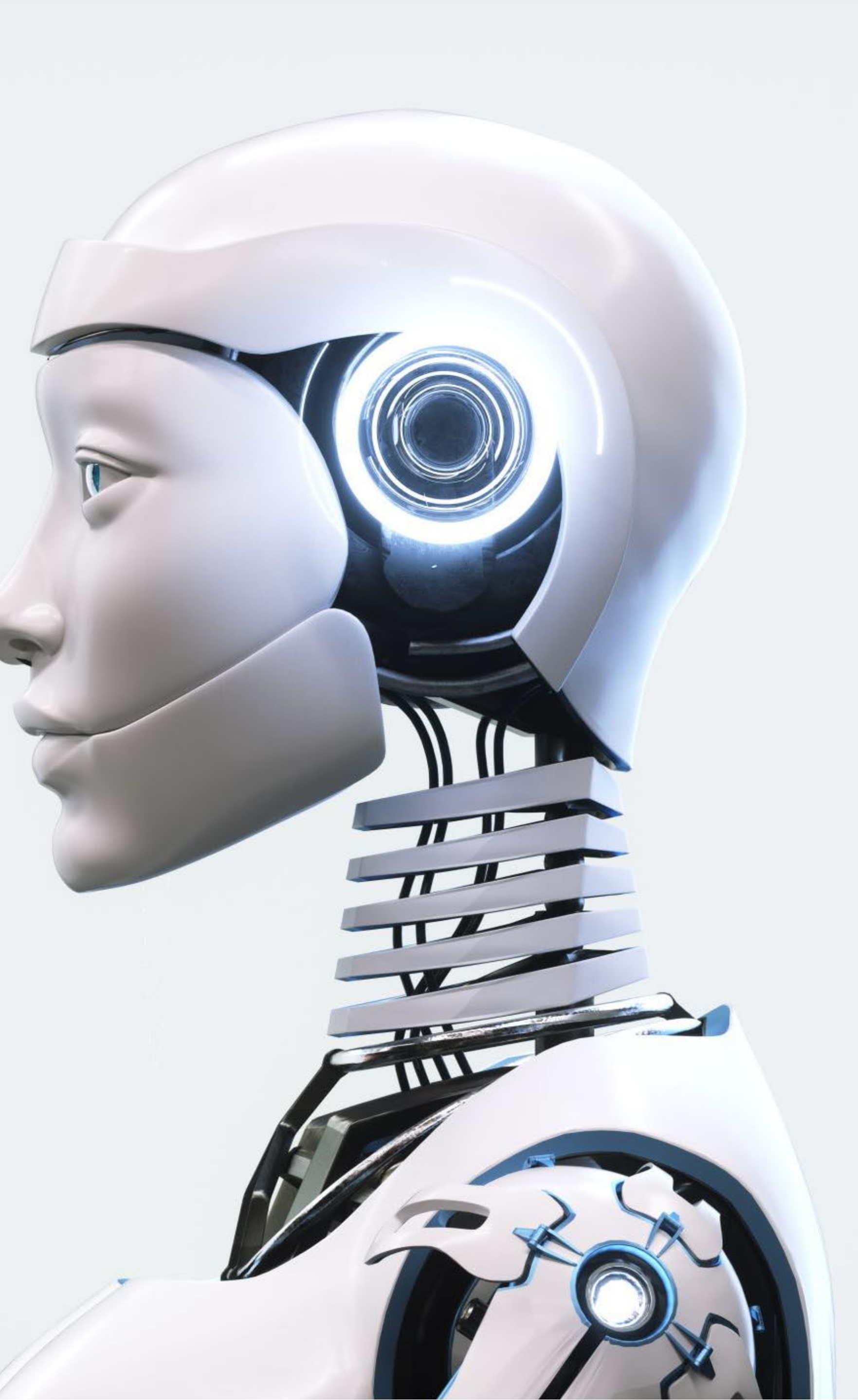
A prognostication

Sam Altman (2023):

“We believe that AI will be about individual empowerment and agency on a scale we’ve never seen before.”

If so, how do we get there from here?

And what are some of the public policy implications?



2023: Year of Generative AI

The Technology

- Generative AI, via Large Language Models (LLMs)
- Other widescale tools – such as small language models (SLMs), multimodal models (MMMs), large action models (LAMs), and cognitive AI – are emerging or on the way.

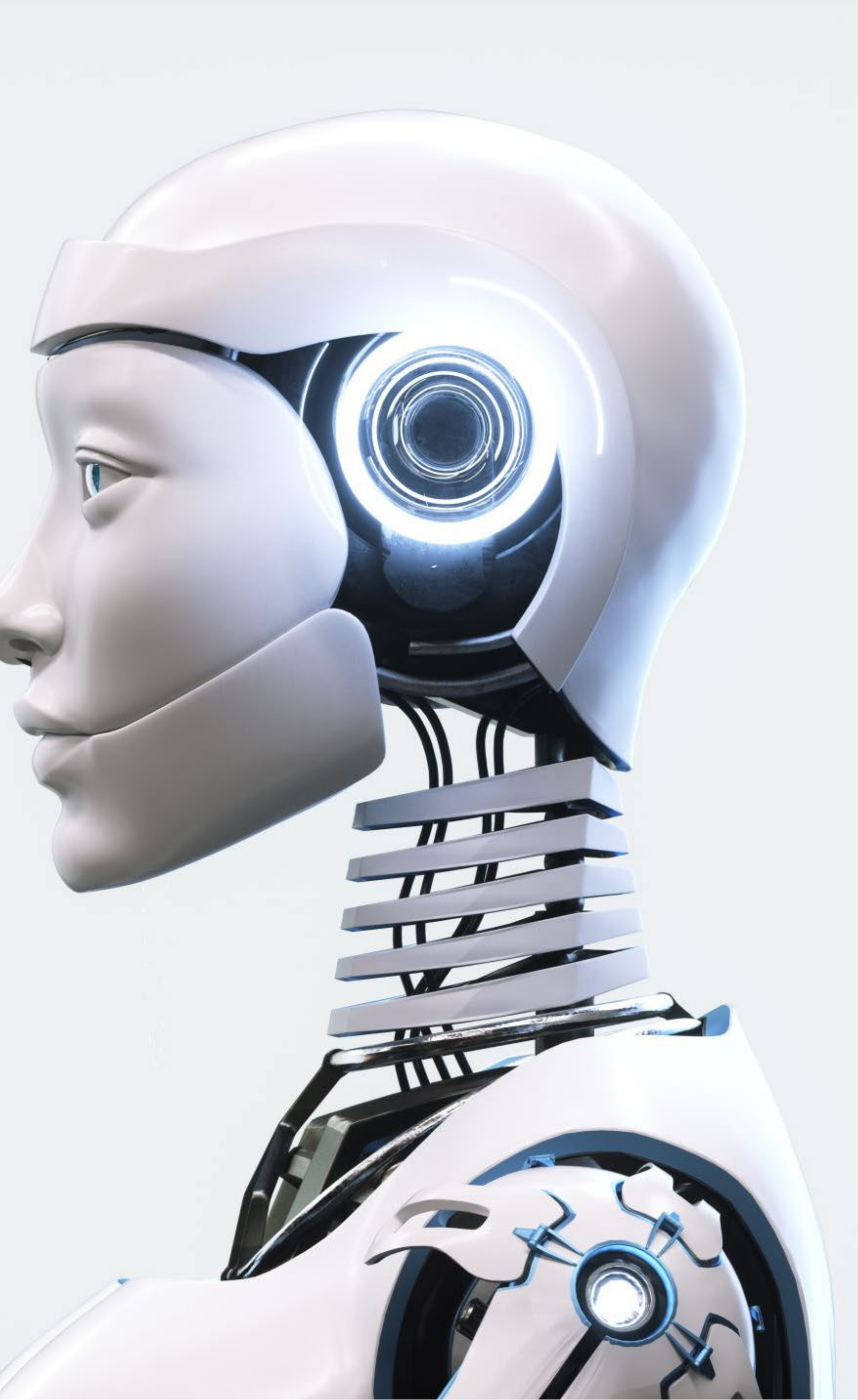
The Key Companies

- OpenAI, Google, Microsoft, Meta, Amazon, Anthropic, others

The Global Public Policy Response

- Substance: Forms of transparency, explainability, bias management, oversight, creating safety “guardrails” by prohibiting the most “risky uses.”

- Actions: EU AI Act, White House EO on AI, NIST AI Risk Management Principles, OECD Principles, Bletchley Park Accord, Hill legislation, etc.



2024: Rise of Personal Digital Agents

- Sam Altman: *"Personal agencies"* are *"going to be a giant thing of the next couple of years,"* that *"a lot of folks are gonna want."*

- Bill Gates: Digital agents will constitute *"a revolution,"* and *"utterly change how we live our lives, online and off…. 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."*

- Web platform companies: OpenAI (GPTs), Google (Bard), Microsoft (Copilot), X (Grok), Amazon (Q, and now Rufus)

- Startups: Personal.ai, Inflection, Adept, Imbue, Lindy, HyperWrite, BabyAGI, more

- Public policy response: ???

Public Policy Agendas

Moving from AI accountability to human agency

Current Policy Agenda: "AI Accountability”

Regulating behaviors of large AI institutions for greater transparency, oversight, and risk management. Building the “guardrails” against harms.

Complementary Policy Agenda: "Human Agency via AI”

Seeding markets with tech and governance inputs to incentivize more competition, innovation, and choice. Building the “merge lanes” for benefits.

A Test Case:

- The “authentic” Personal Digital Agent (PDA)

But what is an authentic Personal Digital Agent?

By definition, being a successful agent involves two interrelated dimensions:

(1) Making decisions and taking actions in the world,

(2) On behalf of someone else

Dimension One: *agenticity*

- OpenAI: An AI system’s “agenticness” is “the degree to which it can adaptably achieve complex goals in complex environments with limited direct supervision.” A measure of capability.

Dimension Two: *agentiality*

- GLIA Foundation: An AI system’s “agentiality” is the degree to which it is actually authorized to ably represent the end user/principal. A measure of relationship.

The Agentic Dimension

Open AI -- Degrees of agenticness for a PDA include:

- Goal complexity
- Environmental complexity
- Adaptability
- Independent execution

Focus should be on mitigating risks and allocating accountability for harms; confirming user-alignment is nascent topic best left for another day.

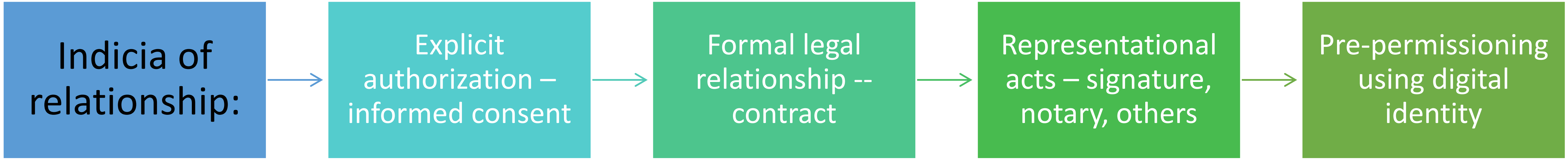
Practices for keeping agentic AI systems safe include evaluating suitability for the task, constraining action-space, setting default behaviors, legibility of actions, monitoring, attributability, and maintaining control.

Source: Yonadav Shavit, Sandhini Agarwal, Miles Brundage, *et al*, *Practices for Governing Agentic AI Systems*, OpenAI White Paper (December 2023).

The Agential Dimension

The "on behalf of" aspect covers sufficiently robust authorization of the individual as principal, including:

- the principal's expectations ("user-alignment") for such representation,
- the agent's base of knowledge about the principal and her intentions,
- acknowledged process for obtaining authorization
- substance of duties and recourse in authorized relationship, and
- tangible indicia of relationship



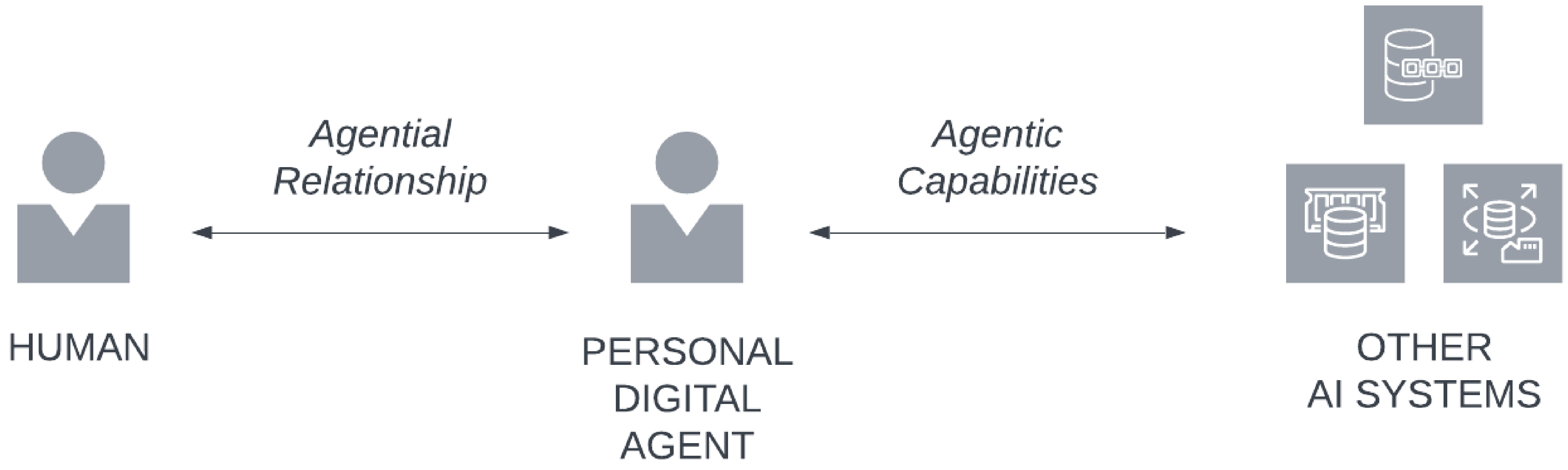
Indicia of relationship:

Explicit authorization – informed consent

Formal legal relationship -- contract

Representational acts – signature, notary, others

Pre-permissioning using digital identity



Two Dimensions, In Action

Enhancing Agentic Capabilities, through AI Interoperability

Interoperability is the glue that holds together disparate systems and networks.

This includes industrial era analog stuff (railroads, telegraph systems, multimodal shipping), and communications/information networks (radio systems, telephone systems, computer accessories, email, the Internet and the Web).

Digital Interoperability

"Interoperators who plug new technologies into legacy platforms help lower switching costs, drive competition and innovation, and give end users real choices." -Cory Doctorow

Digital interoperability is the ability of heterogeneous data-based networks to connect and communicate seamlessly with one another.

Advocates: US FTC, EU Digital Markets Act, New America Foundation, Mozilla, Data Transfer Project/Initiative, proposed US legislation, others

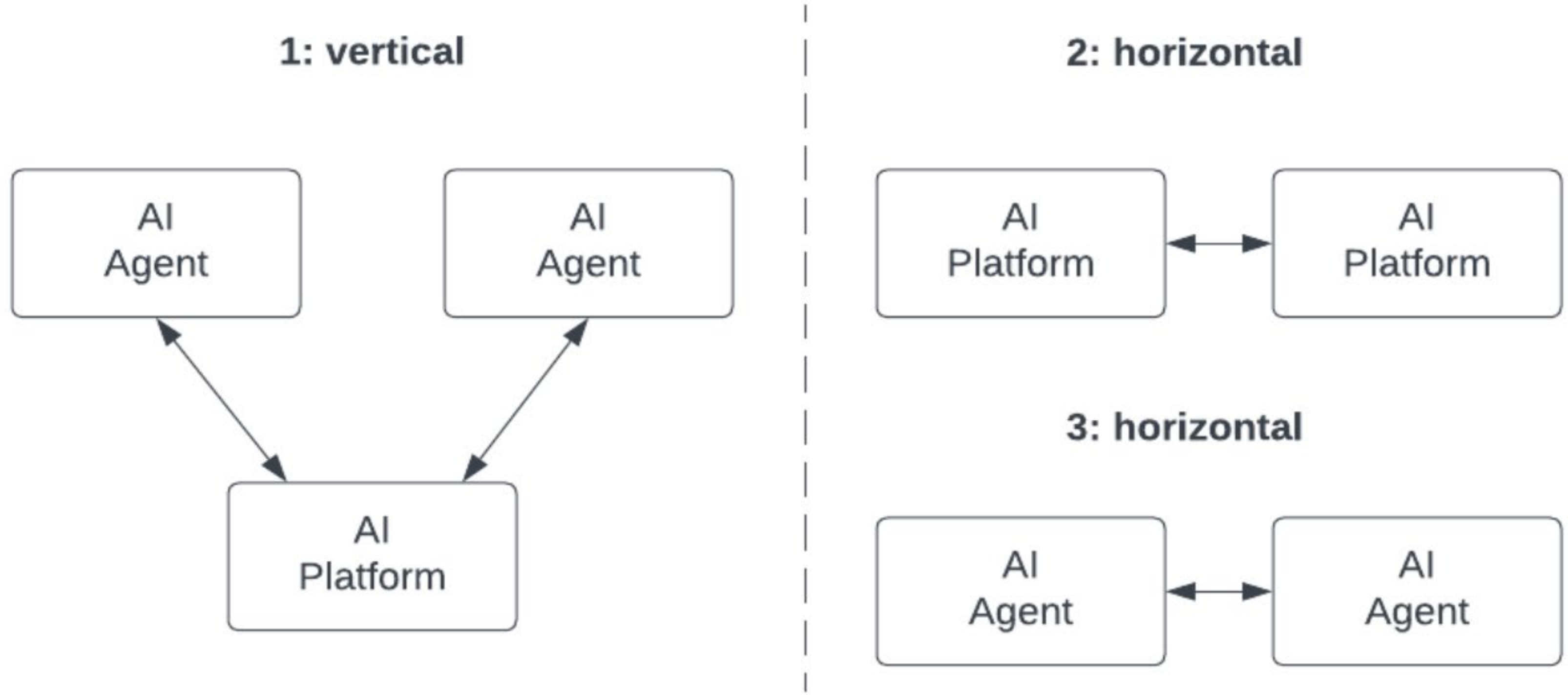
Introducing AI Interop

“AI interop” would allow different AI systems to interact directly, on a seamless, real-time basis.

For the first time, disparate intelligent systems would be able to talk with each other – querying, negotiating, challenging, agreeing – in a whole host of interactions and transactions.

Source: Richard Whitt, *AI Interop: Roots, Benefits Framework, Next Steps*, IEEE Paper (January 2024).

Vertical AI Interoperability



Two possible forms of AI interoperability:

- Horizontal between different players operating in same market
- Vertical between different players in different markets.

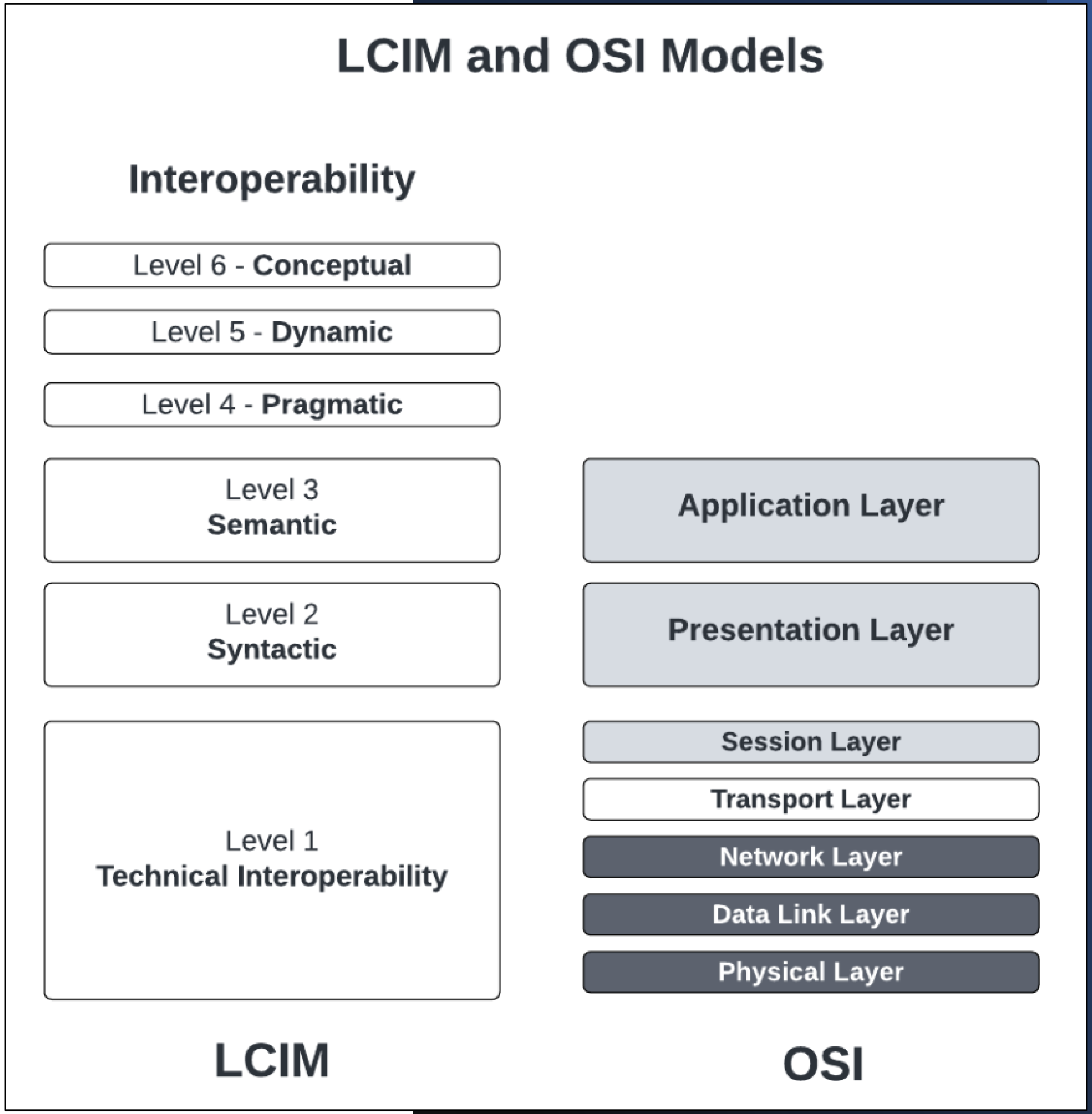
Vertical AI interoperability would allow more seamless interaction between PDAs and larger AI systems and platforms.

Filling the AI Interoperability Gap

Technical: What would constitute an effective interop standard? Can we re-use existing protocols? Are open APIs an adequate substitute/stop-gap?

Governance: Should the interop standard be an open one? Which standards body should take the lead – IEEE, IETF, someone else? What about a role for NIST? For industry groups?

Policy: Is vertical interop an effective remedy for market concentration concerns? What about horizontal interop? Should governments consider inducements? mandates?



LCIM and OSI Models

Interoperability

Level 6 - **Conceptual**

Level 5 - **Dynamic**

Level 4 - **Pragmatic**

Level 3
Semantic

Level 2
Syntactic

Level 1
Technical Interoperability

Application Layer

Presentation Layer

Session Layer

Transport Layer

Network Layer

Data Link Layer

Physical Layer

LCIM

OSI

Enhancing Agential Relationships, through Trusted Intermediation

We should ensure that the robust capabilities unleased by AI interop and other tools actually promote human empowerment and autonomy.

Creating and sustaining authentic relationships with trusted intermediaries, providing us with authentic PDAs, is more a governance challenge than a technical one.

What are some key considerations?

IEEE: “Ethically Aligned Design” endorses personal sovereignty over digital agents

“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 agent is a proactive algorithmic tool honoring their terms and conditions in the digital, virtual, and physical worlds.”

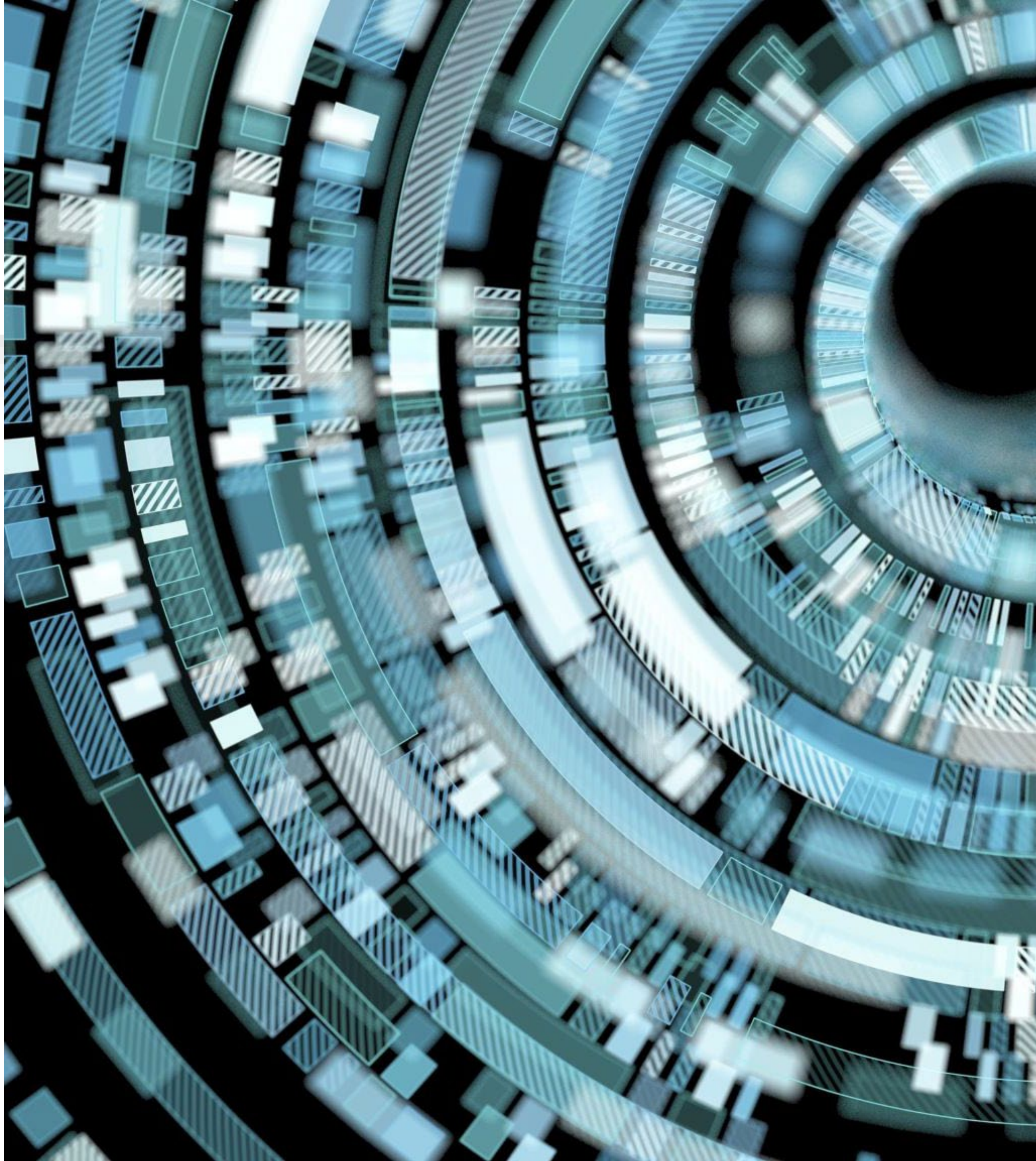
World Economic Forum: Advancing digital agency requires data intermediaries

2022 WEF report highlights human-centric data intermediaries as tangible way to improve trust and valid permissioning in data sharing markets.

Examples:

- Data stewards (e.g., chief data officer)
- Digital fiduciary
- Data collectives: trust, collaborative, cooperative, data commons

Trusted intermediaries also could help individuals/communities outsource their decision-making to AI agents.



One form of agential governance: fiduciaries-based intermediaries

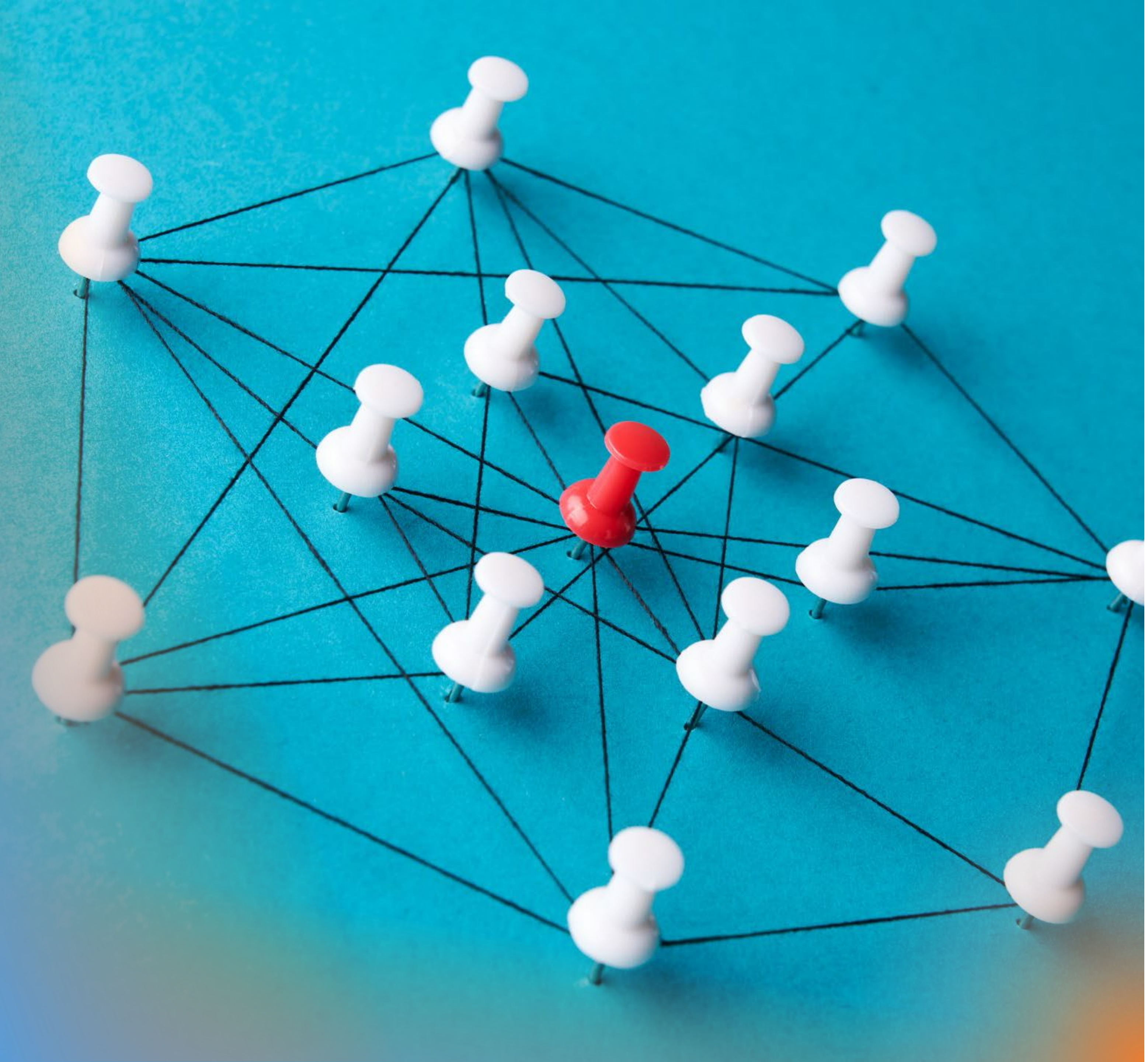
Digital fiduciaries

Data trusts

The common law of fiduciaries and trusts (doctors, lawyers, accountants, librarians) matches well to the 21st Century asymmetric power challenges of Web platforms using personal data and AI.

One proposal is for entities to become trustworthy intermediaries by voluntarily opting into duties of loyalty with their patrons/customers/communities.

Source: Richard Whitt, *Hacking the SEAMs: Elevating Digital Autonomy and Agency for Humans*, 19:1 Colorado Tech Law Journal 135 (2021)

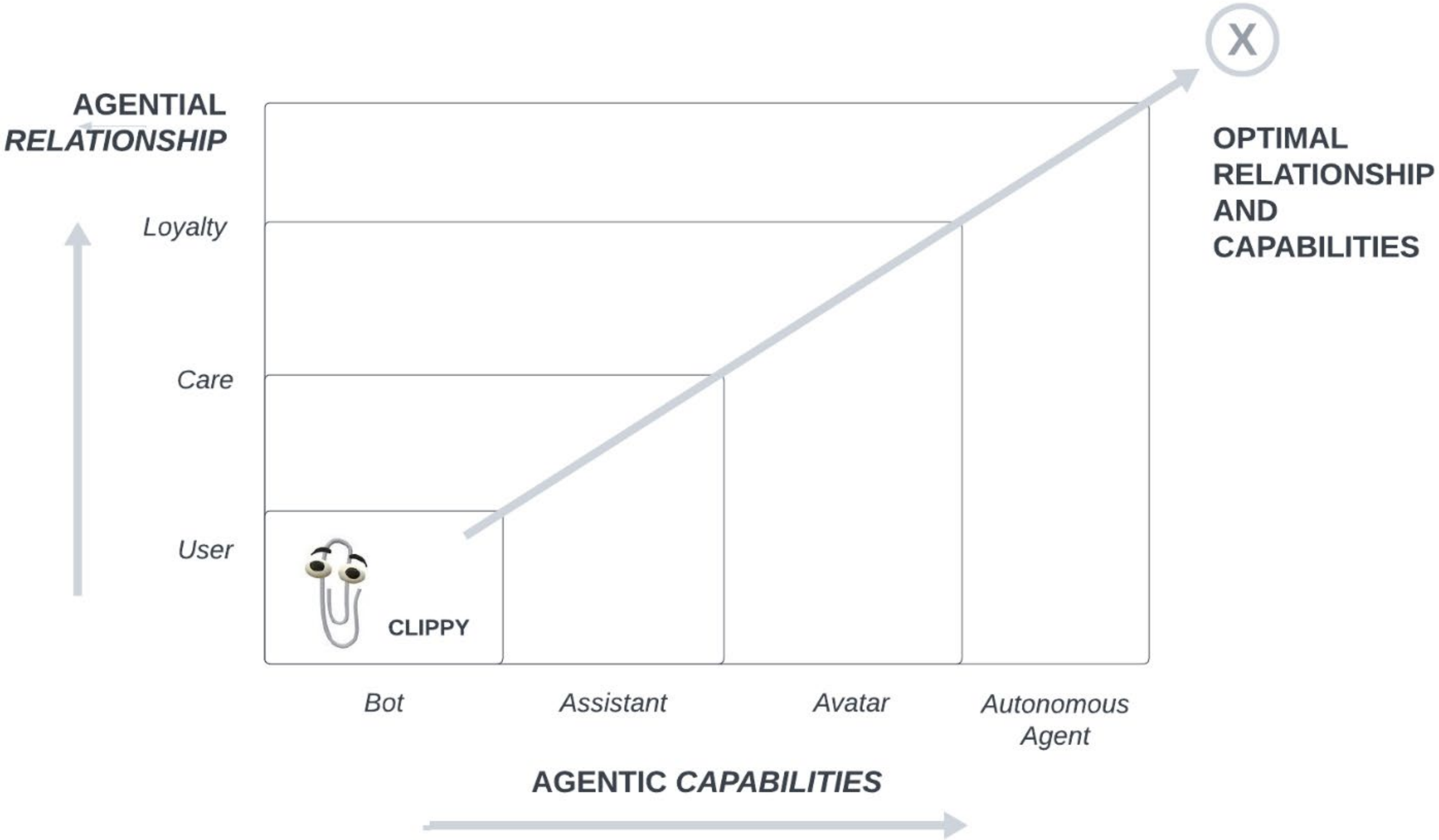


We should want both dimensions of digital agency -- working together.

Spider-Man: *“With great power, comes great responsibility.”*

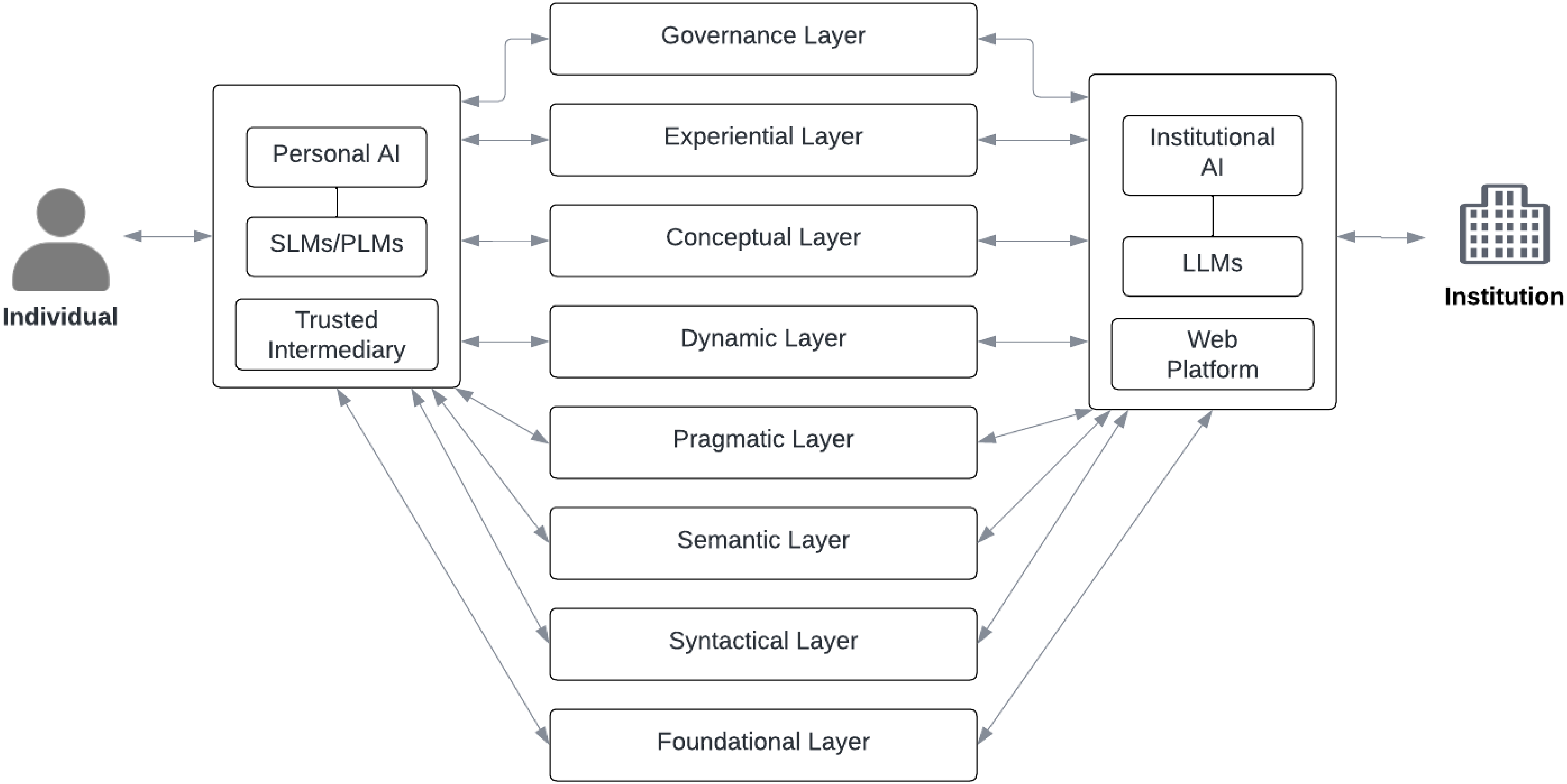
In the AI context, there should be acknowledged, proportional tradeoffs between agenticity (robust capabilities) and agentiality (robust relationship).

Two Digital Agency Dimensions, Interrelated



Some viable use cases for the Authentic PDA

- *Protecting one's personal data flows*
- *Broadcasting one's intentions/policies*
- *Exporting "middleware" of one's filters and decision engines*
- *Circulating one's universal shopping cart*
- *Challenging consequential decision engines*
- *Mediating the terms of one's engagements with Web, virtual, and physical environments*
- *Building our own communities of interest*



A Layered View

Ultimately, we need to develop a new trust-based *ecosystem*

Some possible next steps

OPEN AI INTEROP STANDARDS: ASPIRATIONAL FOR NOW

TRUSTED INTERMEDIARIES: NOW GETTING ORGANIZED

CORPORATE POLICIES: RELYING ON MARKET INCENTIVES

PUBLIC POLICY: NUDGING TECH AND MARKETS TO HONOR DIGITAL AGENCY

A PROPOSED STRAWPERSON: *HUMAN/DIGITAL CODE OF RIGHTS AND DUTIES*

Thank You

richard@glia.net