AI Agents and the Progressive Intermediation Between Users and Online Services: A Critical Analysis

future internet

MDPI

Review

AI Agents and the Progressive Intermediation Between Users and Online Services: A Critical Analysis

Marco Rondina [1,2,*], Antonio Vetrò [1,2,*], Juan Carlos De Martin [1,2], Manuela Bargis [3] and Gabriele Elia [3]

[1] Department of Control and Computer Engineering, Politecnico di Torino, 10129 Turin, Italy; juancarlos.demartin@polito.it
[2] Nexa Center for Internet & Society, 10138 Turin, Italy
[3] TIM S.p.A., 10148 Turin, Italy; manuela.bargis@telecomitalia.it (M.B.); gabriele.elia@telecomitalia.it (G.E.)
* Correspondence: marco.rondina@polito.it (M.R.); antonio.vetro@polito.it (A.V.)

Abstract

The disruptive potential of AI agents is becoming increasingly evident, impacting the very nature of the interactions between users and online services and content. Representing a new form of intermediation, these agents are transforming the way users access and interact with online resources, while also reshaping the distribution of informational power, economic value, and access in the digital ecosystem. This paper analyses the evolution of intermediation mechanisms on the internet through a critical, non-systematic review guided by Critical Systems Heuristics. We trace the progressive layering of intermediation mechanisms to online content/services from the technical infrastructures, and we argue that AI agents constitute a structurally distinct intermediation layer that introduces new dynamics of power, opacity, value distribution, and access to the digital ecosystem, with significant implications for users, online services, and the broader society. We are currently crossing a potentially significant evolutionary shift in the Internet, where AI agent intermediation could progressively exert a substantial influence on how information is exchanged. This paper offers a framework with which to discuss the implications of this potential evolution and to guide future research, public discussion, and policy interventions in this area.

Keywords: internet; digital intermediation; AI agents; power dynamics; information access

1. Introduction

Machine learning and artificial intelligence (ML/AI) technologies have proven their disruptive potential in the last decade. Recent innovations, coupled with the increasing availability of data due to the "computerization of the world" [1] and the advancements in hardware technology, have resulted in their transformation from research niches into the foundation of numerous applications encompassing all aspects of social and economic activities. The Transformer architecture [2] represented a paradigm shift in the potential of the Natural Language Processing (NLP) algorithms, thereby establishing the foundation for Large Language Models [3]. Following the completion of a costly and resource-consuming training phase [4,5] on a massive volume of text data, these models are capable of statistically generating text that appears meaningful, coherent, and syntactically correct. Using these models, numerous chatbots have been developed to facilitate interaction and are specifically designed to simulate interaction with another human being as closely as possible. The high cost of entering this market has led to the predominance of major web entities

Check for updates

Academic Editors: Gianluigi Ferrari and Ivan Serina

Received: 15 June 2026
Revised: 3 August 2026
Accepted: 6 August 2026
Published: 8 August 2026

as the primary owners of such chatbots as Google, Apple, Meta, Amazon, and Microsoft. A limited number of nascent companies were able to compete with the economic power of the GAFAM, such as OpenAI and Anthropic (both of which are highly funded by the GAFAM itself).

The evolution of LLMs has led to the development of highly specialized models that can interact with themselves and with other models, known as agents. This constitutes the defining characteristic of agentic AI, which refers to software capable of executing actions with third parties without the necessity of human intermediation (or very little), except for initial configuration, re-configurations, and human input required to complete a task (e.g., the food a person wish to order), and corresponds to the most advanced of the three levels of agency distinguished in Section 2.

Agentic capabilities could introduce a paradigm shift in the way internet users access online services: a shift that scholars are starting to refer to as the "Agentic Web" [6]. Intermediation in web access is not a completely new phenomenon: at the beginning of the internet era, search engines offered a new, intermediary way of accessing online information and services. However, the statistical and probabilistic nature of LLMs promises much more significant intermediation: rather than searching for a link to access the desired resource, for example, the AI agent provides the most statistically probable answer. In addition, the protocols introduced for this purpose (Section 2) are designed to let such software interact with web services, with other agents, or with other persons. Nevertheless, such technological innovation gives rise to a series of significant implications from various perspectives, including technology, economics, politics, and societal aspect at large.

There is growing scholarly attention to AI agents, such as technical contributions focusing on architectures, applications, and limitations [7–9]. Social science perspectives address specific harms such as work [10] or misinformation [11–13]. Moreover, the transformative potential of agentic technologies is subject to wide-ranging analysis [14]. What is absent is a *systemic* analysis that situates AI agents within the broader trajectory of internet intermediation and examines the implications of the transformation of the structural relationship between users and online services.

This paper addresses this gap with two Research Questions (RQs), by asking the following questions:

(i) *In what ways does AI agent intermediation constitute a structurally distinct layer compared to previous forms of internet intermediation (DNS, search engines, social media platforms, super apps)?*

(ii) *What are the implications of this shift for the distribution of informational power, economic value, and access in the digital ecosystem?*

To answer these questions, this critical, non-systematic review traces the progressive layering of intermediation mechanisms from the Domain Name System to AI agents (Section 4), guided by the Critical Systems Heuristics (Section 3) as an analytical lens to surface distributional consequences and power dynamics that techno-optimistic narratives tend to obscure (Section 5).

2. Background

Intermediation. By definition, intermediation refers to "the act of carrying messages or making connections between people or things that are unwilling or unable to meet" [15]. Across fields like media, finance, innovation, education, and sustainability, intermediation in digital contexts refers to actors and systems that sit in between others to connect, translate, and shape information, value, or relationships. A general definition describes an intermediary as "a person or organization that acts as an agent between other people or things," while digital intermediaries are organizations that bring content from third-party providers to users via digital software, channels, and devices [16,17]. Hutchinson defines digital intermediation as "the combination of digital media actors including technologies, agencies and automation. It is also the deep intricate, and interconnected process of content production, distribution, and consumption which is the result of several unseen infrastructures" [18]. We broadly define intermediation as the process of connecting users to online services and content, and we focus on the progressive layering of intermediation mechanisms in the internet.

Agents. According to the Stanford Encyclopedia of Philosophy, "an agent is a being with the capacity to act, and 'agency' denotes the exercise or manifestation of this capacity" [19]. However, attributing this type of agency to artificial systems is debated in the literature, since it is contested whether the presence of true mental states is required to possess the capacity for agency. Attributing mental states to artificial intelligence systems can reinforce the mystification surrounding AI, presenting it as a kind of magic [20]. This is not our intention, but we continue to use the term "agentic AI" since it is the common way to refer to the tools that we are going to analyze, even if we agree that further studies are needed to find a better and more precise way to identify these technological systems, without reinforcing the ideological statements that contribute to the mystification of this software. Russel and Norvig [21] use a more practical definition of agent based on its inputs and its outputs, i.e., "anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators". They also use this definition to define AI as "the study of agents that receive percepts from the environment and perform actions". Using the same core concepts, Dorri et al. propose a definition that focuses on the fact that the action is driven by the agent's underlying goal, defining it as "an entity which is placed in an environment and senses different parameters that are used to make a decision based on the goal of the entity. The entity performs the necessary action on the environment based on this decision" [22]. The feedback loop is also the foundation of relevant industry definitions, such as the one from IBM, which defines an AI agent as "a system that autonomously performs tasks by designing workflows with available tools" [23]. In this case, the "designing workflows" is intended as the result of the "agentic reasoning" activities. In practice, this implies that the software can conduct a multistep prompt–answer interaction with itself. It is important to note that, in contrast to conventional human reasoning, this interaction is inherently probabilistic in its nature, even when actions are performed and resources interacted with (e.g., an external search engine).

The terms "chatbot", "AI assistant" and "AI agent" are often used interchangeably, although they denote systems with substantially different capabilities. To avoid this ambiguity, we distinguish three cumulative levels along a spectrum of increasing agency, which we use consistently throughout the paper:

1. *Conversational LLM interfaces* (chatbots): Systems that generate responses exclusively from the parametric knowledge acquired during training, with no access to external resources (e.g., the first public release of ChatGPT in 2022);
2. *Answer engines*: Systems that couple an LLM with live information retrieval, synthesizing up-to-date web content into a direct answer that may include source citations (e.g., AI Overviews in Google Search, Perplexity);
3. *Task-executing agents*: Systems that, in addition to the previous capabilities, can autonomously perform actions on third-party services on the user's behalf (e.g., booking, purchasing, submitting information), through tool-invocation and inter-agent protocols such as MCP and A2A (introduced below).

Only the third level fully matches the classical model of intelligent agents proposed by Wooldridge, i.e., software components that are *reactive*, *proactive*, and *social* [24]: task-executing agents do not merely react to prompts, but pursue delegated goals through actions (proactivity) and interact with services and other agents (sociality). In this paper, "AI agent intermediation" refers to the trajectory along this spectrum. Since the deployment of task-executing agents is still at an early stage, the empirical evidence currently available (Section 5) mostly concerns the first two levels; where a claim applies to a specific level of the spectrum, we make this explicit. When we refer to "AI agents" without further qualification, the term encompasses all three levels. This broader usage is warranted by the object of the analysis: what the three levels share is the position they occupy between the user and the source, and the two properties that make that position structurally new are distributed across the spectrum rather than confined to its upper end, since probabilistic generation is already present at the first level and delegated enactment only at the third (Section 4.4). Where a claim concerns a capability rather than that position, we name the level to which it applies.

Protocols enabling agentic interactions. To connect AI applications to external systems, Anthropic introduced a new open-source standard, called Model Context Protocol (MCP): https://modelcontextprotocol.io/docs/getting-started/intro (accessed on 12 October 2025). MCP enables answer engines and task-executing agents to access and interact with various data sources and tools through a standardized interface. The protocol comprises two layers: the data layer and the transport layer. Data is exchanged using a JSON-RPC-based protocol. Transport is supported in two ways: STDIO (Standard Input–Output) for processes on the same machine (e.g., accessing the filesystem), and HTTP for enabling communication with remote servers. The use of the HTTP is mediated by the Representational State Transfer (REST) architectural style, articulated by Roy Fielding [25], paving the way for the programmatic interaction between services that underpin contemporary web applications [26]. HTTP + REST APIs constitute the infrastructure upon which all subsequent forms of internet intermediation are built. Although often invisible to end users, they embody design choices and governance structures that profoundly influence how information and services are accessed, controlled, and monetized online.

Multi-agent architectures and protocols. The natural evolution of single task-executing agents interacting with services and APIs is the growing adoption of multi-agent workflows, in which multiple agents work towards a common goal. Leitão and Karnouskos [27] (Chapter 2.3) identified three main topologies: centralized, hierarchical, and heterarchical. In a *centralized* architecture, there is one entity responsible for all planning tasks and several executors. In a *hierarchical* organization, decision-making responsibility is divided among different nodes based on a pre-defined hierarchical tree. Finally, a *heterarchical* architecture is based on the collaboration and cooperation between nodes, although master–slave relationships between nodes with the same authority are also possible. Händler [28] classified multi-agent frameworks into three categories: *general-purpose systems*, i.e., systems adaptable to a wide range of applications; *central LLM controller*, i.e., a single central control agent that orchestrates different specialized models; *role-agent systems*, i.e., a system where collaboration between agents is based on a conversation-simulated framework. Which of these arrangements prevails is not a neutral engineering choice: it determines where decision-making sits within an agent ecosystem, and therefore who controls the conditions under which intermediation operates. In the context of multi-agent interaction, the de facto standard is the Agent2Agent (A2A) protocol. It enables different AI agents to communicate and collaborate as peers. Originally developed by Google and subsequently donated to the Linux Foundation in 2024, A2A emerged from the recognition that agent-to-agent communication requires fundamentally different architectural assumptions than agent-to-tool interactions: https://a2a-protocol.org/latest/topics/what-is-a2a/ (accessed on 13 October 2025). A2A operates through a client–server architecture in which AI agents function as both clients and servers, exposing HTTP endpoints that implement standardized communication primitives. The protocol's core abstraction is the Agent Card, a JSON metadata document analogous to a digital business card that describes an agent's capabilities. Communication between agents typically utilizes the same HTTP + JSON/REST architecture presented for MCP. This allows agents to collaborate over the long term, building on previous interactions and preserving intermediate results, which makes it easier to track and manage overall progress towards a complex goal.

3. Methodological Stance

3.1. Scope and Sources

This paper takes the form of a critical non-systematic review in the sense of Grant & Booth [29]: it synthesizes and critically analyses evidence from diverse sources rather than providing an exhaustive or systematic mapping of the literature. To compare successive layers of internet intermediation, it develops six structural dimensions and an explicit criterion for distinguishing differences of degree from differences of kind, set out under *Answering RQ1* below.It deliberately focuses on risks and distributional harms, because these are systematically underrepresented in the predominantly techno-optimistic discourse on agentic AI, and because identifying them is a precondition for adequate safeguards and governance interventions.

Within this genre, source selection was purposeful [30]. It drew on the logic of *critical case* sampling: rather than a representative survey, it prioritizes pivotal cases whose prominence and resources make their exposure diagnostic, so that a distributional harm observed even in well-established, well-resourced instances of a phenomenon can reasonably be expected to affect more fragile ones. The cases discussed in Section 5 are offered as evidence that a distributional dynamic occurs and is consequential, not as a measure of its prevalence, and selection reflects relevance to the dimensions of implication rather than an exhaustive account of the field. This selectivity is a deliberate feature of the critical stance adopted here, whose limitations we make explicit in Section 6.

Evidence running against the argument was handled in two ways, both visible in the text. For RQ1, the structural criterion was applied to every layer rather than to AI agents alone. Where it returns continuity instead of novelty, this stated in Section 4.4. For RQ2, where a countervailing development bears on a specific dimension, it is reported there, as in Section 5.3. The remaining ones are collected in Section 6, together with the concurrent factors that our causal account cannot fully disentangle.

3.2. Critical Systems Heuristics

CSH is a framework designed to support reflective practice in technology assessment and systems design [31,32]. It provides a structured approach to examining the boundary judgments that define who is included or excluded from a system's design, whose interests are privileged, which are the premises of existence of a given form of technology, and which consequences are considered legitimate versus externalized. In the scientific literature, CSH has been applied to different contexts (including technology), and "is most frequently applied to address (...) power asymmetries" [33], which is the goal of our work.

Our discussion is conducted from the perspective of Requirements Engineering, which is useful to consider how the different perspectives of stakeholders affect architectural design choices. This perspective also determines which boundary questions the analysis focuses on, as specified below. The twelve questions of CSH are organized into four sources of influence: (1) the sources of motivation, which concern the purpose of the system and the interests it serves; (2) the sources of control, which concern who has decision-making authority and how it is exercised; (3) the sources of knowledge, which concern whose expertise is embedded in the system and how it is validated; and (4) the sources of legitimacy, which concern who bears the costs and risks of the system and how they are recognized. Table 1 summarizes the CSH questions. How the framework is applied to each research question is specified in the two paragraphs that follow.

Table 1. The twelve boundary questions of Critical Systems Heuristics (CSH) [32], organized by source of influence, social role, and concern, following the Requirements Engineering approach [31].

Sources of Influence	**Social Roles (Stakeholders)**	**Specific Concerns (Stakes)**	**Key Problems (Stakeholding Issues)**
Motivation	(1) Who is/ought to be the intended beneficiary of S?	(2) What is/ought to be the purpose of S?	(3) What is/ought to be S's measure of success or improvement?
Control	(4) Who is/ought to be the decision maker in control of the conditions of success for S?	(5) What resources and conditions of success are/ought to be under the control of S's decision maker?	(6) What conditions of success are/ought to be outside the control of S's decision maker (the decision environment)?
Knowledge	(7) Who is/ought to be providing the relevant knowledge and skills for S?	(8) What are/ought to be the relevant expertise, knowledge and skills necessary for the operation of S?	(9) What are/ought to be the assurances of successful implementation of S, i.e., what is its guarantor?
Legitimacy	(10) Who is/ought to be the witness representing the interests of those negatively affected by S but not involved with S?	(11) What are/ought to be the opportunities for emancipation of those negatively affected to have expression and freedom in the worldview of S?	(12) What space is/ought to be available for reconciling different worldviews regarding S among those affected and involved?

3.3. Answering RQ1

To address RQ1, we conduct a comparative analysis of the five intermediation layers examined in Section 4: DNS, search engines, social media platforms, super apps, and AI agents. Each layer is treated as a sociotechnical system, examining both its technical operation and the power relations embedded in its design; the AI agent layer, as the object of this paper, is additionally interrogated through four of the CSH boundary questions, one from each source of influence (Section 5). The comparison is organized along six structural dimensions:

- *User navigational agency*: The extent to which the selection among alternative sources or services remains an explicit step of the interaction, performed by the user. It instantiates the CSH sources of **control**: who decides which options are available and which one is enacted;
- *Source attribution*: The extent to which the origin of what is presented to the user is identifiable and independently verifiable. It instantiates the sources of **knowledge**: whose knowledge the system conveys and what guarantees its validity;
- *Interaction modality*: The technical form of the exchange between the user and the intermediary;
- *Epistemic status of the output*: Whether what the user receives is retrieved from existing, identifiable artifacts or probabilistically generated. Together with interaction modality, it is a descriptive property of the exchange rather than an instantiation of a single CSH source, although it bears directly on the CSH question of the guarantor (Table 1, question 9), i.e., what assures the validity of what the system produces;

- *Value extraction logic*: The mechanism through which the intermediary captures economic value from the interactions it mediates and from the content produced by third parties. It instantiates the sources of motivation: whose interests the system serves and by which measure of success;
- *Governance model*: Who defines and enforces the rules of the layer, and through which accountability mechanisms. It instantiates the sources of **legitimacy**: how those affected by the system are represented in its governance.

The choice of this set is itself a boundary judgment, which we make explicit through two requirements. Jointly, the dimensions must cover the constitutive elements of the intermediation relation, as defined in Section 2: who selects, what is known about the source, how the exchange takes place, what kind of object the output is, who captures its value, and who sets its rules. Individually, each dimension must be comparable across the trajectory; where a layer exists before the property that a dimension captures, this is stated explicitly ("not applicable") rather than forced into an artificial value. The set is not exhaustive: other candidate dimensions, such as the degree of personalization of the output, overlap substantially with the six adopted here and are discussed within them where relevant. To distinguish differences of degree from differences of kind, we adopt an explicit criterion: a layer is *structurally* distinct if, on at least one dimension, it introduces a configuration that was absent from all previous layers, rather than an intensification of an existing one. The criterion is applied to every layer in the sequence, so that the identification of each of them is tested rather than assumed; DNS, being the first layer considered, is assessed against direct, unmediated addressing by IP.

3.4. Answering RQ2

To address RQ2, we examine the distributional consequences of AI agent intermediation across six dimensions of implication: traffic redirection (Section 5.1), economic value (Section 5.2), geopolitical power (Section 5.3), social effects (Section 5.4), cultural production (Section 5.5) and telecommunications infrastructure (Section 5.6). The dimensions were identified inductively, through an exploratory reading of the phenomenon, which made it possible to identify the areas where structural consequences are already observable or can be credibly predicted; the boundary questions were then applied deductively to interrogate them.

Of the twelve boundary questions (Table 1), we address four in detail, one for each source of influence: who benefits (Q1), who controls the conditions of success (Q4), whose knowledge the system draws upon (Q7), and who represents those affected but not involved (Q10). These four constitute the distributional skeleton of the analysis, namely who gains, who decides, who supplies, and who bears. They are the *social roles* questions of the CSH matrix, one for each source of influence: consistent with the Requirements Engineering perspective adopted here, the analysis is scoped to who occupies each role, leaving the stakes and stakeholding issues addressed by the remaining questions outside its scope. Each is answered in the *is*/*ought* form that distinguishes CSH from a thematic checklist, and each dimension in Section 5 opens by stating the questions it primarily bears on. The comparison is conducted for agent intermediation as a general configuration, not for a specific deployment.

The evidence base was assembled in two phases, between September 2025 and June 2026. An exploratory phase identified the dimensions above; a targeted phase then gathered evidence for each, combining peer-reviewed literature with gray literature, including industry reports, policy documents, journalistic investigations and technical publications. Gray literature is included because peer-reviewed evidence on a rapidly evolving phenomenon structurally lags behind it and often provides the first documented observations. How a source is used depends on whether it reports a primary observation or an account of one, on how mature the evidence it presents is, and on what interest its author or organization has in the claim being made. If a claim is based on preliminary journalistic reporting, or describes an observed practice rather than a measured effect, this is stated alongside the claim. If a primary document exists, such as an incident report, an official statement or an industry report, it is cited directly rather than relied on through secondary coverage. Where a source has a stake in the prospect it describes, as with industry forecasts on network evolution, this is stated as well.

Because the object of the analysis is the intermediation between user and information rather than generative AI in general, evidence concerning general model behavior is admitted only where its relevance to the intermediation function is made explicit, and claims are anchored to the level of the intermediation spectrum (Section 2) to which they apply. Evidence concerning conversational interfaces and answer engines is used to ground extrapolations about task-executing agents, and is marked as such.

4. RQ1: The Progressive Intermediation Between Users and Online Services

This Section traces the evolution of internet intermediation to provide context for the impact of AI agents' intermediation. The contemporary internet is characterized by multiple, overlapping forms of intermediation, each of which introduces different mechanisms of control, value extraction and power concentration. This section traces the evolution from basic technical infrastructure (Section 4.1) to information-oriented search engines (Section 4.2), social media platforms and super apps (Section 4.3) and, finally, AI agents (Section 4.4). This progression reveals how intermediation has shifted from simply facilitating access to actively shaping what users see, how they interact, and which services they can access. Each layer builds upon and transforms the previous ones, creating increasingly opaque and centralized systems of control that set the scene for exploring the potential implications of AI agent intermediation. Figure 1 summarizes the progression of the intermediation between user and content.

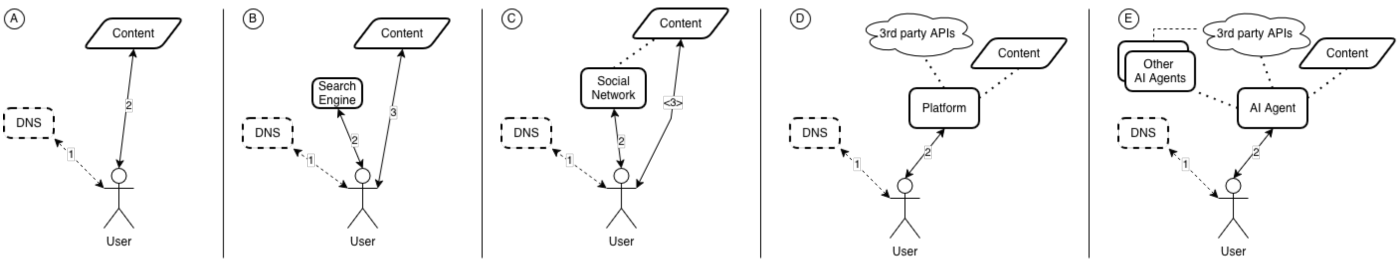

Figure 1. The progressive intermediation between users and online services. Panels (**A**–**E**) correspond to the five successive layers of intermediation examined in this paper: (**A**) the basic technical intermediation of the Domain Name System (DNS); (**B**) search engines; (**C**) social media platforms; (**D**) super apps; and (**E**) AI agents. The numbers mark the ordered steps of the user's path to the content: (1) resolution of the domain name through the DNS, common to every layer; (2) direct retrieval of the content in panel (**A**), and the exchange with the intermediary in panels (**B**–**E**); and (3) the further step to the source, which the user performs in panel (**B**), performs only occasionally in panel (**C**), and no longer performs in panels (**D**,**E**), where the intermediary reaches the content and any third-party services on the user's behalf (dotted arrows).

4.1. Basic Technical Intermediation: The Domain Name System

The World Wide Web, introduced by Tim Berners-Lee in 1989 at CERN, was designed as a decentralized system for sharing information across networked computers [34]. The HyperText Transfer Protocol (HTTP), formalized through RFC 1945 in 1996 [35], established the foundation for client–server communication on the web. HTTP's request–response model enabled the retrieval and display of hypertext documents, creating the familiar browsing experience. In its earliest form, accessing web resources required direct knowledge of Internet Protocol (IP) addresses, the numerical identifiers that uniquely locate devices on a network. This approach imposed significant cognitive burdens on users who needed to memorize or maintain lists of numeric addresses to access desired resources. The Domain Name System (DNS), standardized first in 1983 through RFC 882 and RFC 883 [36], introduced the first layer of technical intermediation. Other technical mechanisms, such as directory services (e.g., LDAP), also perform forms of intermediation; we take DNS as the entry point of the user-facing path to online content and services, rather than attempting an exhaustive account of intermediation at the network and infrastructure layers by translating human-readable domain names (such as example.com) into machine-readable IP addresses (Figure 1A). From the outset, DNS has been used alongside other, previous internet services, such as FTP (1971) and Telnet (1971). The date refers to the first RFC for a proposed TELNET protocol RFC (97) and email (1981), and its core functionality has been instrumental in the diffusion of the World Wide Web. The hierarchical, distributed naming system effectively abstracted the underlying network complexity, enabling users to navigate the internet using memorable text-based identifiers rather than numeric strings [37]. While DNS is often perceived as purely technical infrastructure, it represents a critical intermediation layer that shapes internet governance and access control. The management of top-level domains and the infrastructure of DNS root servers concentrate significant power in the hands of specific organizations, primarily ICANN (Internet Corporation for Assigned Names and Numbers) and a limited number of entities that operate root name servers [38–40]. Moreover, DNS has been a focal point for censorship and control, as governments and corporations can manipulate DNS records to block access to specific websites or services [41,42]. A recent example of the relevance of DNS services in the modern internet is the global unavailability of Amazon Web Services on 19 October 2025, when a DNS error made AWS unreachable worldwide. The official summary of the incident written by AWS [43] says: "The incident was triggered by a latent defect within the service's automated DNS management system that caused endpoint resolution failures for DynamoDB" [43,44]. While DNS was designed to facilitate access to websites, it does not solve the problem of information discovery, which became increasingly apparent as the web expanded.

4.2. Search Engines Intermediation

As the web expanded exponentially throughout the 1990s, the limitations of directory-based navigation systems became apparent. Early attempts at organizing web content, such as Yahoo's manually curated directory (The Yahoo! Directory is available on the Internet Web Archive: https://web.archive.org/web/20141122194515/https://dir.yahoo.com/, accessed on 23 October 2025; more web directories can be found on: https://en.wikipedia.org/wiki/List_of_web_directories, accessed on 23 October 2025), could not scale with the rapid growth of online information. Search engines emerged as a solution to this information discovery problem, introducing a fundamentally new form of intermediation between users and web content.

Google, launched in 1998, revolutionized web search through its *PageRank* algorithm, which ranked pages based on their link structure rather than simply matching keywords [45]. This approach transformed web search into a sophisticated intermediation service that interpreted user intent and curated access to information (Figure 1). Google's dominance in search, commanding over 90% of global search engine market share by the mid-2010s [46], established it as the primary gateway to the internet for billions of users, influencing them [47]. The cognitive and behavioral impacts of search engine intermediation have been documented in psychological and information science research. The "Google effect" [48] shows that people are significantly less likely to remember information that they know can be easily retrieved online. Subsequent research has revealed that frequent use of search engines can affect not only memory formation but also metacognitive judgments, leading users to overestimate their own knowledge [49].

Beyond cognitive impacts, search engines have reshaped the business models underlying web services [50]. The auction-based model for search advertisements [51], where advertisers bid for placement alongside relevant queries, introduced economic incentives that fundamentally altered content creation and web optimization practices, giving rise to the entire industry of Search Engine Optimization (SEO) [52]. The tracking mechanisms that personalize the results of a search engine query as much as possible have paved the way for enormous privacy concerns [50,53].

The political and economic power concentrated in search engines has drawn increasing attention. Search engines function as gatekeepers that determine which information gains visibility and which remains obscured. The opacity of ranking algorithms and their susceptibility to manipulation raise concerns about democratic access to information and the potential for censorship, whether imposed by governments or exercised through corporate policy [54]. Recent antitrust investigations in multiple jurisdictions have examined whether dominant search engines abuse their market position to favor their own services and disadvantage competitors [55].

While ostensibly providing neutral tools for finding information, search engines actively shape what information is seen, by whom, and under what circumstances. This intermediation goes far beyond technical facilitation, constructing the informational landscape through algorithmic curation that reflects the interests, biases, and business models of their operators.

4.3. Platform Intermediation, Social Media, and Super Apps

While search engines mediate access to information, platform intermediation represents a more comprehensive form of control over digital interactions, services, and economic transactions. Digital platforms emerged to intermediate interactions between multiple user groups, reshaping how individuals access and engage with online content and services. Plantin et al. describe and explore the differences between "infrastructures" and "platforms" (see Table 1 in [56]). For example, they identify the focal interests of the two as public value and private profit. Platform intermediation encompasses two distinct layers of the trajectory, treated in turn below: social media platforms and super apps, which differ in how value is extracted and in how access to third-party services is governed.

Social media platforms. Social media platforms exemplify this shift toward platform-mediated internet access (Figure 1). By 2024, Meta's family of apps (Facebook, Instagram, WhatsApp) reached over 3 billion daily active users globally [57], making these platforms the primary interface through which billions of people access news, communicate with others, and discover online content. The phenomenon of social media platforms becoming the "front door" to the internet became particularly evident, for example, in the controversy surrounding news distribution. Platforms like Facebook altered news consumption patterns, with significant portions of users accessing news primarily through social media feeds rather than visiting news websites directly [58]. This shift raised concerns about the sustainability of journalism, as platforms captured advertising revenue, while news organizations bore the costs of content production. Moreover, the algorithmic curation of news feeds, optimized for engagement rather than informational value, has been implicated in the spread of misinformation and the creation of filter bubbles that reinforce existing beliefs [59].

Super apps. The platform intermediation model has reached its most comprehensive form in the form of "super apps", i.e., platforms that extend beyond their primary focus (usually messaging or social networking) to create a digital ecosystem of new functionalities, either developed directly by the platform owner or by third parties (Figure 1). WeChat in China is one example. Launched in 2011, WeChat evolved beyond messaging to integrate social networking, mobile payments, e-commerce, government services, and countless third-party mini-programs within a single ecosystem [60]. This "app-within-an-app" architecture creates an enclosed digital environment where users conduct virtually all their online activities without leaving the platform. While super apps offer convenience through integration, they also concentrate unprecedented control over digital life in the hands of single entities, raising concerns about surveillance, data exploitation, and the erosion of open web principles. The economic structure underpinning platform intermediation relies heavily on what Zuboff terms "surveillance capitalism" [53]. Platforms extract value by collecting vast amounts of behavioral data, which they use to predict and influence user behavior, primarily for targeted advertising. This business model creates perverse incentives that prioritize engagement over user welfare, contributing to addiction-like usage patterns [61] and to mental health impacts particularly among young users [62].

Platform power extends to encompass significant political influence. Platforms act as private regulators of speech, making content moderation decisions that affect billions of users with limited transparency or accountability [63]. The ability to de-platform individuals or organizations, to algorithmically amplify or suppress content, and to shape the information environment gives platforms gatekeeping power that rivals that of traditional media institutions. The concentration of platform power has prompted regulatory responses worldwide. The European Union's Digital Services Act [64] and Digital Markets Act [65] aim to impose obligations on large platforms regarding content moderation, data portability, and interoperability.

4.4. AI Agents Intermediation

The widespread diffusion and use of chatbots, answer engines, and task-executing agents introduces a qualitatively different type of intermediation, building upon and transforming previous layers. Unlike search engines, which return lists of links that users must select and navigate to access the source websites, and unlike social platforms, which typically display content attributed to identifiable authors, AI agents generate direct answers by statistically synthesizing their training data or by performing actions to fulfill the request in the prompt using their agentic capabilities (Figure 1). The potential actions needed to complete a task are enacted by RESTful HTTP APIs, with which the task-executing agent can interact using the Model Context Protocol (MCP) introduced in Section 2.

Where DNS transparently maps domain names to IP addresses, search engines visibly rank websites with metadata indicating their source (even if the ranking algorithms are not fully transparent, since they are influenced by advertising and editorial control), and social platforms identify content creators through profiles, AI agents present synthesized responses in which the relationship to the original sources is opaque or entirely absent. The natural language interface enables users to formulate complex (and ambiguous) queries without needing to optimize keywords for algorithmic interpretation, as required by search engines. However, this conversational accessibility comes with a fundamental trade-off: the probabilistic nature of text generation means that outputs may contain plausible but factually incorrect information, a phenomenon documented extensively in the literature on hallucinations [66]. Unlike search results, where users can evaluate source credibility before engaging with content, or social media posts, where author identity provides signals for trust assessment, agents' responses combine information from multiple sources to create coherent and syntactically correct prose without enabling verification of the underlying data. This represents the most complete form of informational intermediation yet deployed at scale: where previous layers mediated access to information while preserving some degree of source transparency, AI agents' intermediation synthesizes multiple sources to create a single, seemingly authoritative answer, fundamentally altering the user's relationship to information verification and source attribution.

Table 2 summarizes the comparison of the intermediation layers across the different dimensions and allows us to answer RQ1 explicitly. The criterion is applied to every layer, not only to AI agents: each column marks the dimensions on which that layer introduces a configuration absent from all its predecessors. For AI agents, new configurations appear on five dimensions (source attribution, interaction modality, epistemic status, value extraction and navigational agency), while only the governance model is a reconfiguration of degree. These five do not, however, amount to five independent fractures: they follow from two structural novelties, each of which reconfigures several facets of the intermediation relation.

Table 2. Comparative matrix of the five intermediation layers along the six structural dimensions defined in Section 3. For each layer, **[K]** marks a dimension on which it introduces a configuration absent from all preceding layers; for the AI agent layer, the dimensions on which the change is only one of degree are marked [D].

	Technical (DNS)	**Search Engines**	**Social Media**	**Super Apps**	**AI Agents**
User navigational agency	The user specifies the destination, the system resolves it	**[K]** Choice among ranked alternatives determined by the intermediary	**[K]** Selection within an algorithmically curated feed, without an expressed query	**[K]** Choice among services admitted to the ecosystem	**[K]** The intermediary enacts the selection; the user no longer chooses among presented alternatives, which remain optional provenance
Source attribution	Not applicable	**[K]** Results identify their sources; ranking criteria opaque	Content tied to author profiles; amplification logic opaque	Services identified within platform boundaries	**[K]** Citations are possible only with external retrieval; a response drawn solely from parametric knowledge has no discrete source to attribute
Interaction modality	**[K]** Name resolution by a third party	Keyword query returning a ranked list	Feed browsing	In-platform execution of services (mini-programs, payments)	**[K]** Delegation of a goal in natural language, rather than a discrete query or command
Epistemic status of the output	Deterministic resolution of an existing address	Retrieval of existing documents	Retrieval of existing user content	Execution of predefined services	**[K]** Probabilistic generation: output synthesized anew, need not correspond to any existing source, and unverified
Value extraction logic	Administrative registration fees	**[K]** Advertising auctions on query intent; behavioral data	Advertising on engagement; user-generated content and data	**[K]** Commissions on transactions and payments, in addition to advertising and data	**[K]** Subscriptions and usage fees; third-party content absorbed at training time is monetized with no traffic or attribution to producers

Table 2. *Cont.*

	Technical (DNS)	Search Engines	Social Media	Super Apps	AI Agents
Governance model	**[K]** Multi-stakeholder institutions (ICANN, IETF); documented public processes	**[K]** Unilateral corporate control of ranking; ex-post antitrust scrutiny	Private content governance; emerging statutory obligations (DSA/DMA)	**[K]** Corporate gatekeeping of market access for third-party services	[D] Provider-defined model behavior; dedicated accountability regimes only nascent (e.g., EU AI Act implementation)

The first novelty is **the epistemic status of the output**. Every previous layer surfaces artifacts that already exist: DNS resolves an address, search engines and social feeds retrieve and rank existing documents and posts, super apps execute predefined services. AI agents instead generate their output probabilistically, synthesizing text that need not correspond to any existing source and that is returned without a human verification step. This is not a more opaque form of retrieval but a different operation, and it appears already at the first level of the spectrum defined in Section 2, where even a conversational interface with no external access generates rather than retrieves. Because the output is generated rather than retrieved, source attribution and value extraction are reconfigured in turn: a response drawn from parametric knowledge has no discrete source to which it can be attributed, and the third-party content that trained the model is absorbed into its weights and monetized without any traffic or attribution returned to its producers.

The second novelty is that **the intermediary takes over the enactment of the interaction**, and it reconfigures two further dimensions. On interaction modality, the user delegates a goal in natural language rather than issuing a query or making a selection. In the case of a navigational agency, the final selection performed by the user in each previous layer, however limited the set of choices had become, is passed on to the intermediary. The response or transaction is produced rather than being chosen from among the presented alternatives. That an answer engine or a task-executing agent may still display citations or links does not restore this act, since these are provenance appended to an answer already produced, not a set of sources the user must choose among to reach the content.

The two novelties span the spectrum outlined in Section 2: probabilistic generation is present already in the simplest conversational LLM interface, whereas delegated enactment appears only in the task-executing agent.

On the one remaining dimension, the table records continuity: the governance of AI agents stays within the unilateral corporate configuration established by earlier platforms, extended rather than remade. We note that when a search engine itself returns a generated answer, as with AI overviews, it is operating as an answer engine in the sense of Section 2 and thus already belongs to the layer under analysis, rather than constituting a counterexample to it. Platform curation is sometimes offered as an equivalent case, since feeds and platforms both assemble content and retain the user. The difference lies on the epistemic axis: a feed selects and orders items that continue to exist independently and remain attributed to their authors, whereas an agent produces text that may correspond to no existing item and to no identifiable author. The same distinction locates a further analogy, sometimes drawn with earlier news aggregators such as Google News: an aggregator likewise captures value from content it does not produce, but it surfaces existing articles under their authors' names and links back to them, whereas an agent's response does not necessarily correspond to a retrievable item or an identifiable author.

The comparison does not rest on a count of marked dimensions. Search engines, as the first layer to mediate content rather than addresses, also introduce new configurations on four dimensions, but on the epistemic axis they, like every pre-agent layer, retrieve what already exists. What singles out AI agents is not the number of reconfigured dimensions but the two structural novelties from which those reconfigurations follow: a new kind of object as output and the transfer of enactment to the intermediary. Because the criterion requires only a single such dimension, the claim of structural distinctiveness is secured by either of them.

5. RQ2: Implications of AI Agent Intermediation

The diffusion of AI agents represents a fundamental transformation in the architecture of digital intermediation between users and online service providers, carrying potentially profound implications for the distribution of digital traffic, economic value, and informational power. Before examining the six dimensions of implication individually, we read them through four CSH boundary questions, one from each source of influence, contrasting the current state of AI agent intermediation (*the is*) with a normative reference (*the ought*). These four questions (i.e., who benefits, who controls the system, whose knowledge it draws upon, and who represents those affected) organize the distributional analysis that follows. Table 3 sets out the four answers. Read together, they describe a single asymmetry rather than four separate findings: in each case, the actors who supply the knowledge on which the system depends, or who bear its costs, are distinct from those who benefit from it and those who decide how it operates. The six dimensions examined in the remainder of this section provide the evidence on which these answers rest.

Table 3. The four CSH boundary questions used to structure the analysis of implications, one for each source of influence (Table 1), contrasting the current configuration of AI agent intermediation (*is*) with a normative reference (*ought*).

Boundary Question	Is	Ought
Who benefits? (*motivation, Q1*)	The corporations that own the underlying models and infrastructure capture the traffic, data, and economic value previously distributed across content producers and platforms	Benefits should also accrue to the users, content creators and knowledge communities whose contributions make the systems possible, rather than being concentrated in a handful of providers
Who controls the conditions of success? (*control, Q4*)	A small number of US- and China-based corporations define model behavior, deployment and access, retaining the capacity to restrict or deny it	A layer that mediates access to information and services should be subject to accountable, public-interest governance rather than to unilateral corporate control
Whose knowledge does the system draw upon? (*knowledge, Q7*)	Content creators, commons-based communities and data workers supply the knowledge, and their contributions are absorbed into the models without attribution or compensation	These providers should be recognized, attributed and compensated, so that their continued contribution remains sustainable
Who represents those affected but not involved? (*legitimacy, Q10*)	Content creators losing revenue, users subject to manipulation, linguistic and cultural communities underserved by the models, and data workers in precarious conditions have no structural representation in the system's design	Their interests should be given standing through safeguards and accountability mechanisms; in their absence, this analysis is offered as one such witnessing perspective

5.1. The Systematic Redirection of Internet Traffic

In the terms of Table 3, this dimension bears on whose knowledge the system draws upon (Q7) and who benefits from it (Q1): the communities that produced the corpus on which agents rely are those from which the system withdraws visibility. This dimension is specific to the intermediation function: the redirection follows from the agent occupying the position between the user and the source, and would not arise from a model used as a standalone tool. The most immediate and measurable effect of widespread AI agent adoption is the **systematic redirection of internet traffic** away from original content sources toward the platforms hosting the deployment of the underlying Large

Language Models. This phenomenon is not just a change in user navigation patterns, but a structural transformation in how value flows through the internet ecosystem. The disruption to agent intermediation presents a different aspect to that seen with the advent of content-based platforms such as YouTube or Spotify: here, the intermediary is not hosting the content, but rather synthesizing it from multiple sources. When users query a conversational interface or an answer engine rather than visiting a website directly, the agent provider captures both the user inputs and the desired (but not delivered) content, while the original content creators—whose works comprised the training data enabling the agent's responses—potentially receive neither traffic nor attribution. This dynamic is already empirically observable across prominent knowledge platforms. Stack Overflow, the collaborative programming Q&A platform that has served as a critical resource for software developers since 2008, reported significant traffic declines following ChatGPT's launch in November 2022 [67]. The platform's traffic, which had remained relatively stable throughout the 2010s, experienced measurable reductions as developers increasingly turned to conversational interfaces for coding assistance and troubleshooting guidance. The same displacement now extends to task-executing agents that operate directly on a codebase, resolving the developer's problem without any consultation of the platform; we report this as an observed shift in practice rather than as a measured effect. Similarly, Wikipedia, the collaboratively edited encyclopedia that represents one of the internet's most successful *commons-based* peer production initiatives, has attributed declining traffic to the rise of conversational interfaces or answer engines that synthesize information from Wikipedia articles without directing users to the source [68,69]. Because Stack Overflow and Wikipedia are robust on precisely the dimensions that ordinarily sustain a knowledge commons—their scale, funding, prominence, and editorial oversight—a decline measurable even in them is significant rather than incidental: where the most established communities already register the effect, more fragile ones, with fewer contributors and thinner resources, can reasonably be expected to experience it at least as acutely. The impact is particularly severe for Wikipedia editions in languages with smaller contributor communities, where declining traffic risks creating what researchers have characterized as a "doom spiral": reduced visibility leads to fewer new contributors, which, in turn, diminishes content quality and further accelerates traffic decline [70]. In his Journalism and Technology Trends and Predictions 2026 report [71], the Reuters Institute clearly states the concerns related to the paradigm shift of search engines becoming "answer engines": "Publishers expect traffic from search engines to almost halve (−43%) over the next three years, following recent dramatic declines in referrals from social media. [...] A number of studies show that the proportion of 'Zero-click searches'-where the user doesn't click through to any website-increases significantly when these [AI] overviews are shown". In their analysis, they noticed that from November 2024 to November 2025, Google traffic from organic search to 2.500 websites under analysis dropped by about 30% globally.

The quantitative evidence reported in this subsection was produced against conversational interfaces and answer engines, the first two levels of the spectrum defined in Section 2. It bears on task-executing agents because the mechanism is the same and operates more completely at that level: an agent that completes a task on the user's behalf removes the visit to the source no less than one that answers in its place.

These cases illustrate a fundamental tension: **the AI systems that now compete with these platforms for user attention were trained extensively on the very content these communities produced**. Stack Overflow's corpus of developer Q&A and Wikipedia's encyclopedic articles constitute precisely the high-quality, structured knowledge that makes LLMs capable of providing useful responses in technical and factual domains. Yet the economic and social benefits that previously accrued to these platforms through user visits, community engagement, and contributor recognition are systematically captured by the corporations controlling the AI models. This traffic redirection extends beyond individual platforms to represent a broader shift in Internet architecture. Where previous forms of intermediation—search engines and social platforms—still directed users to content sources (even as they inserted themselves as powerful gatekeepers), AI agents reduce, and for many queries remove, the need for users to visit original sources. Where this happens, the synthesized response becomes the endpoint of user interaction, making source websites invisible to the user. This architectural change threatens the sustainability of the open knowledge commons that underpins much of the internet's informational infrastructure, as content creators and platform communities lose both the traffic-based revenue (through advertising) and the community engagement that motivates continued contribution.

5.2. The Economic Consequences

In the terms of Table 3, this dimension bears on who benefits (Q1) and on those who bear the costs without being represented (Q10): compensating the people who produce content and label data lies outside the system's own measure of success.

The disruption of compensation follows directly from the redirection described above, and the intellectual property tension is specific to intermediation in a further respect: the synthesized output substitutes for the work at the moment of access, not only in the corpus used for training. The conditions of data labor, by contrast, are a property of how these models are produced rather than of how they mediate; they are reported here because they answer the question of whose knowledge and labor the intermediary draws upon (Q7).

The economic implications of this traffic redirection are particularly acute when considering the disruption of established monetization models. The contemporary internet economy has been predominantly sustained by advertising revenue, which fundamentally depends on user attention and engagement metrics derived from website visits. This status quo has prevailed at the expense of alternative ways of managing the internet infosphere, such as the commons, cooperative platforms, participatory technologies, convivial technologies and so on. Content creators have relied on this advertising-based model to finance their production costs. When AI agents intercept user queries and provide synthesized responses without directing traffic to the original sources, **the link between content creation and economic compensation is effectively broken**. This disconnection creates a paradox: AI models depend on high-quality content for their training data and continued relevance, yet their very operation undermines the economic mechanisms that enable such content production. AI agents can also influence another important aspect of the internet economy: the online sale of goods and products. AI agents can redirect user searches and queries towards websites, e-commerce platforms or sellers that pay the model owner, or have signed an agreement with them, similarly to what happens with search engines, but on a more subliminal level. The insertion of advertisements in LLM responses, including content not clearly marked as advertising, is already a research topic [72].

A related but distinct set of issues concerns intellectual property. While **AI agents synthesize information from copyrighted materials to generate their responses, the legal status of these synthetic outputs is a controversy**. A notable illustration of these tensions emerged in the creative industries, where Studio Ghibli, the renowned Japanese animation studio, found its distinctive artistic style being replicated by AI image generation models trained on copyrighted works without authorization or compensation [73]. This case exemplifies a broader challenge: AI systems can absorb and reproduce stylistic and substantive elements of copyrighted works in ways that existing intellectual property frameworks struggle to address. The synthesis and transformation performed by AI models occupy an uncertain legal space between derivative works (which typically require permission from copyright holders) and new creations (which might constitute fair use or fall outside copyright protection entirely). This ambiguity creates significant challenges for content creators seeking to protect their work and maintain viable economic models, while simultaneously enabling AI companies to extract value from creators' labor without adequate compensation or recognition. A study by the European Union Intellectual Property Office on the development of generative AI from a copyright perspective [74] highlights that the development of AI systems from a copyright perspective is currently being shaped by ongoing litigation between rights holders and GenAI providers.

The progressive intermediation between users and services through AI agents opens different questions on the impacts on the **labor market**. The more dominant intermediation becomes, the greater the demand for additional capabilities. In this process, the work of data workers is fundamental, since these systems cannot function without human effort being invested in collecting, labeling, and processing data [75], even though AI agents are starting to be deployed within the annotation pipeline itself [76]. Furthermore, many systems rely on intensive data extraction from platform users, who are encouraged to produce data by the offer of free services [53,77]. This "underground" work, which often takes place in precarious, underpaid (or even unpaid) and stressful conditions, breaks the narrative that portrays artificial intelligence systems as a fully automated solution. Crawford and Joler [78] identify such dynamics as new forms of exploitation, where the benefits of AI innovation accrue primarily to wealthy nations and powerful corporations, while the environmental and social costs are disproportionately borne by communities in the Global South. The data labeling work essential for training AI systems is similarly concentrated in lower-income regions, where workers perform cognitively and psychologically demanding tasks under precarious conditions for minimal compensation. This global division of labor in AI production mirrors historical patterns of colonial extraction, adapted to digital technologies.

5.3. Geopolitical Power Concentration

In the terms of Table 3, this dimension bears on who controls the conditions of success of the system (Q4): the resources on which agent intermediation depends are concentrated in the same actors that can restrict access to it.

Concentration in the market for models is a feature of the AI industry at large; what is specific to intermediation is that the same actors control the layer through which users reach information and services, which converts market position into gatekeeping over access. The material dependencies discussed below, in specialized hardware, energy and water, are preconditions for operating this layer rather than consequences of its intermediation function, and are reported as such.

The centrality of AI agents in the relationship between users and information reproduces and amplifies pre-existing patterns of **technological, economic, and power concentration**. The high cost of developing and training LLMs has resulted in market dominance, in the frontier segment, by a limited number of major technology corporations, primarily based in the United States and China. Google, Apple, Meta, Amazon, Microsoft, DeepSeek, OpenAI and Anthropic control the fundamental infrastructure upon which AI agents operate. This oligopolistic structure raises fundamental questions about access, control and the capacity of these corporations to shape the evolution of digital intermediation according to their strategic and commercial interests. An example is the export control directive issued in June 2026 by the US government "to suspend all access to Fable 5 and Mythos 5 by any foreign national, whether inside or outside the United States, including foreign national Anthropic employees". More details can be found in Anthropic's official statement: https://www.anthropic.com/news/fable-mythos-access (accessed on 14 June

2026). This concentration is a tendency rather than a fixed outcome: open-weight models and local deployment, as well as the growing number of smaller providers, mean that more people can access an agent. Open standards for tool use and inter-agent communication also lower the barrier to entry. While these developments may distribute provision more broadly, they do not alter the intermediary position between user and source that the analysis concerns: they change who may occupy it, not whether it is occupied. Rather than just a tool for productivity, AI risks entrenching existing socioeconomic hierarchies by allowing corporations to control the means of production. This centralization allows entities to optimize algorithms for profit at the expense of fairness, potentially disadvantaging workers and marginalized groups [79].

The concentration of control over AI models translates directly into **control over data flows and user interactions**. When organizations or individuals use AI agents powered by third-party models in their day-to-day internet activities, they might share, under contractual terms, sensitive information with the model providers. This creates asymmetric power relationships where the providers of the underlying models gain unprecedented visibility into patterns of use, queries, and behaviors across vast user populations. The privacy implications extend beyond individual data protection concerns to encompass questions of collective data governance and the potential for surveillance at scale [53]. Furthermore, model providers possess the technical capacity to restrict or deny access to their services, whether for commercial, political, or regulatory reasons. This gatekeeper power enables selective exclusion that can have significant consequences for individuals, organizations, and even entire nations who find themselves dependent on AI infrastructure controlled by foreign corporations [80].

As with all web server architectures, these systems require adequate hardware, especially on the server side, as well as network hardware to enable connection between the two. This is particularly relevant for AI agents, which **require more computational resources than previous layers of intermediation** such as search engines [81]. The production of AI systems is firmly rooted in extractive industries, highly specialized industries, and energy-intensive computing infrastructure. The concentration of raw materials (e.g., rare earth elements) creates strategic dependencies that can be leveraged for geopolitical purposes. Similarly, the manufacturing of specialized hardware (e.g., GPUs, TPUs) is dominated by a few companies and countries, creating bottlenecks and vulnerabilities in the supply chain. The AI Index 2026 Annual Report from the Institute for Human-Centered AI of the Stanford University states that "NVIDIA AI chips currently account for over 60% of total compute, with Google and Amazon supplying much of the remainder and Huawei holding a small but growing share" [82] (p. 29) and that "TSMC is a single point of dependency in the global AI supply chain, as it fabricates virtually every leading AI chip, including NVIDIA's Blackwell GPUs and AMD's MI300X" [82] (p. 31). Finally, energy infrastructure becomes a critical resource [5] in determining which nations and regions can sustain large-scale AI deployment. The water consumption associated with cooling data centers adds another dimension to these resource dependencies [83].

The infrastructural dependencies and control asymmetries, paired with the centralization of data flows, acquire their most dangerous manifestation in the accelerating **militarization** of AI technologies. The deployment of AI weapons in contemporary conflicts indicates that this is not merely a hypothetical future scenario but is already emerging, according to early journalistic reports, in the Russo–Ukrainian war [84], in the 2026 Iran conflict [85] and in the Israeli military's Lavender system [86]. All these systems, but especially the latter one, are designed to address "the human bottleneck for both locating (...) and (...) approving the targets", exploit, and aggregate different data sources. Where the delegation of enactment that characterizes agent intermediation extends to systems of this kind, the concentration of control described above would bear on decisions whose consequences are irreversible. All these aspects raise the question of how governance frameworks can ensure that the benefits of agent intermediation are equitably distributed, rather than concentrated in the hands of those who control the underlying infrastructure.

5.4. Social Implications

In the terms of Table 3, this dimension bears on who represents those affected but not involved (Q10) and on who decides (Q4): the users exposed to unequal access, persuasion, and moderation take no part in the decisions producing those effects. The effects discussed here are documented for generative systems in general, and specifically for conversational interfaces; the intermediation function changes their reach, because a system that is the user's principal route to information leaves no visible alternatives against which its framing could be checked. The discussion about power dynamics also extends to **equality of opportunity** and the potential for AI agents to **exacerbate existing inequalities**. Unlike technical intermediaries, search engines, social media platforms, and super apps, the performance of the AI agents is directly related to the size of the underlying LLMs, which, in turn, are related to the amount of data and computational resources that can be invested in their training. This means that the advantage each individual and community can gain from using this technology, where agents come to mediate access to information and services, is directly related to the amount of money that they can invest to access the best performing agents, i.e., the ones that have a larger context window, the ones that have been trained on more data, the ones that have been fine-tuned for specific tasks, etc. Here the risk is to create a new digital divide, an *AI divide*, in which the inequalities between those who can afford to access the best performing agents and those who cannot, are exacerbated.

AI is also affecting the **income gap**. Although artificial intelligence systems have only recently become widespread, the uneven effects they have on economic development are already being felt. These differences are based on both collective and individual factors. At a collective level, these tools have a greater impact in areas and regions characterized by advanced infrastructure and a strong industrial and digital focus. This suggests that more advanced regions may have greater development opportunities than less advanced regions, thus widening the gap. At an individual level, differences in background and professional experience are significant, as the impact of AI on the job market is not uniform across all sectors. Zhang et al. examined the Chinese context from this perspective and found that AI development worsens the income gap [87].

The role of AI agents as the main point of access to information has profound consequences in terms of **political power and democracy**. Research on AI propaganda shows that even relatively modest language models can consistently hold ideological positions across different conversational contexts, producing persona-driven political messaging [88]. This includes intentional strategies to affect the "decision-making capacities of groups, effectively shaping their ability to take advantage of information in their environment" [89]. AI agents create opportunities for manipulation at unprecedented scale, where automated systems can persuade voters in ways that differ from, and are more subliminal than, traditional political advertisement [90]. Their use as content moderation systems could increase opacity, complicate fairness and justice in sociotechnical systems, and obscure the political nature of mass moderation [63]. Without serious regulatory intervention, the flexibility in determining content removal criteria could lead to established censorship systems [91].

The amplification of **social polarization** represents the other side of the coin. Research on personalized AI outputs demonstrates that when LLMs tailor their responses based on users' political affiliations, they generate content that reinforces partisan divisions. Studies examining ChatGPT's outputs when prompted with political context found that left-leaning users received more positive statements about left-leaning political figures and media outlets, while right-leaning users encountered more favorable portrayals of right-leaning entities [92]. This pattern of algorithmic personalization recreates the filter bubble dynamics previously observed on social media platforms, where users encounter information ecosystems that reinforce their existing worldviews while excluding opposing viewpoints. When AI agents serve as the main way of accessing information and providing support for decision-making, these filter bubbles become more prevalent and harder to spot than content curated by platforms, as users may perceive AI-generated responses as objective summaries rather than outputs filtered by perspective. The apparent paradox is that the same AI systems could be used to polarize ideologies and improve engagement on social media (and therefore be profitable), or to influence collective thinking within predetermined "guardrails", depending on who controls their deployment and configuration. Research on narrative polarization demonstrates that different LLMs encode divergent ideological perspectives in their outputs, with measurable differences in how they frame issues related to liberalism, conservatism, and philosophical approaches to justice and governance [93]. As Maximilian Kasy highlights, the intrinsic power problem of AI is that the optimization goals that drive model behavior are defined by those who own and control it [94].

5.5. Cultural Consequences

In the terms of Table 3, this dimension bears on whose knowledge is embedded in the system (Q7) and on who speaks for those it marginalizes (Q10): the linguistic and cultural perspectives thinly represented in the training data are equally absent from the design. Synthetic content and model degradation are properties of generative systems as such; the intermediation function supplies the vector, since agents feed synthesized material back into the same commons whose traffic they divert (Section 5.1). Their cultural defaults, likewise, propagate at the point of access because the agent operates as a universal interface. The cultural implications of this transformation reveal how AI agents are beginning to function not merely as tools but as active participants in cultural evolution, capable of generating and propagating narratives and contents that compete and affect human cultural production [95]. The most immediate cultural consequence is **the growth of synthetic content generated by AI systems**. This phenomenon represents a fundamental shift in the composition of the digital information ecosystem. AI models now produce vast quantities of text, images, and video that populate online spaces, blurring the boundaries between human-created and machine-generated content. This creates a feedback loop whereby AI systems are increasingly trained on outputs produced by other AI systems, raising concerns about model collapse, degradation of output quality and the progressive detachment of AI-generated content from grounded human experience and knowledge [96]. Illustrative examples of this phenomenon have emerged in collaborative knowledge production platforms. Reports indicate instances where Wikipedia articles, particularly in less-monitored language editions, have been substantially authored or edited by AI systems, with contributors potentially using LLM-generated text without clear disclosure [70,97]. While such practices may increase the quantity of available content, **they can fundamentally alter the epistemological character of these resources**, transforming them from repositories of human knowledge verified through collaborative peer review into assemblages of statistically plausible text whose relationship to factual accuracy is probabilistic.

Where previous intermediation layers at least preserved some trace of authorship and provenance, **AI agents synthesize information** from multiple sources into seamless outputs that obscure their origins. This architectural opacity makes it increasingly difficult for users to **assess the reliability of information** they encounter. The absence of visible attribution mechanisms eliminates the social and institutional signals that have traditionally enabled readers to calibrate their trust in information sources, creating an environment where fake news, propaganda, and deliberately misleading content become increasingly difficult to identify. The challenge is amplified by deployment in conversational interfaces, as users may anthropomorphize the system and trust it, creating new vulnerabilities to manipulation.

AI agents also raise questions about **cultural homogenization** [98]. The training process for LLMs involves vast but ultimately finite datasets, predominantly sourced from English-language content and reflecting the cultural perspectives, values, and norms of Western, educated, industrialized, rich, and democratic (WEIRD) societies [4,99]. When these models are deployed globally as universal interfaces for information access and content generation, they risk imposing linguistic and cultural uniformity that marginalizes minority languages and alternative ways of understanding the world. The economic and technical dominance of a few model providers means that the operational constraints (such as model architectures, training datasets, and deployment policies) are established by a few corporate decisions. The risk is not only that AI agents will inadequately represent cultural diversity, but also that they will actively reshape cultural practices towards the convergence of norms and assumptions embedded in dominant AI systems.

The **cognitive implications** of widespread adoption of AI agents encompass fundamental changes in the way humans process knowledge. Research on the "Google effect" demonstrated that reliance on search engines diminishes people's ability to remember information they know can be easily retrieved online [48]. AI agents can amplify this externalization of memory by reducing, or even eliminating, the need to navigate to information sources: users need only to articulate their questions in natural language, and the agent provides synthesized answers. Preliminary research suggests that overreliance on chatbots or answer engines for cognitive tasks may lead to atrophy of skills related to problem-solving, analytical thinking, and creative ideation, particularly among younger users who have never developed these capabilities without AI systems [100–102].

These cultural consequences are interconnected dimensions of a broader transformation in the production and transmission of knowledge. However, unlike previous transitions, the AI transformation is occurring with unprecedented rapidity, compressing adaptation timescales of individuals, institutions and societies.

5.6. The Impacts on the Future of the Telecommunication Sector

In the terms of Table 3, this dimension bears on who controls the infrastructure (Q4) and on who represents those subject to it (Q10): network-level decisions delegated to agents affect all traffic traversing them, including that of users with no visibility into those decisions. Unlike the preceding dimensions, this one concerns the effects of the intermediation layer downwards, on the infrastructure that carries it, rather than on the relation between users and services; it is included because the traffic and processing demands of agent intermediation reshape the network on which all other layers depend.

Task-executing agents' diffusion has the potential to alter the actual **traffic patterns**, since they typically play a dual role, acting both as a client and as a server. This requires low-latency communication channels (especially for real-time applications [103]) with enough bandwidth to support multi-modal data types (text, files, audio, video) [104], but also extensive intermediate processing steps, context sharing, and state synchronization. When deployed at scale, agents engage in constant communication with cloud-based models and other agents within multi-agent systems, as well as with various external services and APIs. The resulting demand for uplink traffic could put additional strain on network architectures that are optimized for consumer internet usage.

This shift in traffic patterns is likely to become more pronounced as technologies relating to the **next generation of telecommunications networks (6G)** become more widespread. In fact, industry forecasts suggest that the desired capabilities of 6G technologies can only be achieved through close integration with advanced ML/AI systems [105]. The **6G *sensing* technology** is designed to enable the network itself to function as a distributed sensor array by analysing disturbances in radio signals. This capability would enable the real-time mapping of the physical environment, potentially eliminating the need for dedicated sensor deployments. This would fundamentally transform the network from a communication channel into a perception system [106], capable of introducing AI-driven network optimizations. The stream of environmental information generated through 6G sensing would have to be processed, interpreted, and acted upon in real time, introducing **new infrastructural requirements**, such as the dynamic allocation of the computational resources necessary for AI inference at strategically positioned edge and cloud locations [107].

The technical capabilities associated with 6G sensing and AI-driven network optimization are frequently presented in industry discourse as enabling **hyper-personalization** of services, whereby telecommunications networks would continuously adapt to individual user behaviors, preferences, and contextual conditions [106]. Conversely, the reliability of this type of network management will need to be carefully scrutinized, as relying on the output of probabilistic software can lead to specific errors and inconsistencies. From a personal rights perspective, the narrative of hyper-personalization obscures the fact that the technical infrastructure required to enable such personalization would be fundamentally indistinguishable from a **surveillance apparatus**. Unlike previous network generations, which only tracked explicit user actions such as website visits or application usage, this surveillance would operate at a more granular level, extending to ambient environmental sensing and more detailed physical behavioral patterns.

The evolution of telecommunication networks also presents challenges in terms of **cybersecurity**. Networks exploiting autonomous agents to manage a 6G infrastructure would expand the attack surface to multiple nodes [108]. Each agent (e.g., one agent monitoring traffic patterns, another optimizing routing decisions, and a third managing resource allocation) would become a potential vector for injecting malicious instructions into other agents' reasoning processes. Adversaries could manipulate the network-level data on which AI agents rely upon for decision-making. This would poison the agents' understanding of network conditions, causing them to make systematically flawed optimization decisions [109]. Where a compromised web application affects only its direct users, a compromised agent operating within network infrastructure could impact all traffic traversing that infrastructure. At the same time, AI systems offer the potential to **enhance threat detection** by analysing large volumes of network traffic data and identifying anomalous patterns. This could enable more proactive security responses [110,111]. The risks here lie in deploying AI-based security systems to protect AI-based network management, creating recursive dependencies if both the network management agents and the security monitoring agents are vulnerable to prompt injection or adversarial attacks. As a sector of essential infrastructure, security failures in AI-based network management could have consequences that extend far beyond the commercial interests of individual telecommunications operators.

Table 4 provides a structured overview of the six dimensions of implication examined in this section, together with the key findings identified for each.

https://doi.org/10.3390/fi18080419

Table 4. Summary of the main dimensions and key findings from the analysis of AI agent intermediation implications (Section 5).

Dimension	Key Findings
Traffic redirection (Section 5.1)	• AI agents provide synthesized answers, eliminating the need to visit original source websites; • Stack Overflow and Wikipedia report measurable traffic declines; • Publishers expect organic search traffic to fall; • Zero-click searches increase when AI overviews are displayed; • Threatens the sustainability of open knowledge commons.
Economic consequences (Section 5.2)	• Advertising-based monetization disrupted: traffic no longer reaches content producers; • Redirection of product queries to paying commercial partners; • Unresolved intellectual property disputes: training on copyrighted material without authorization or compensation; • Rise of data work in precarious conditions.
Geopolitical power (Section 5.3)	• Market dominated by a handful of US and Chinese corporations; • Providers gain unprecedented visibility into user data flows; • Gatekeeper power: restriction of access to AI infrastructure; • Strategic dependencies on rare earth elements, specialized hardware (GPUs/TPUs), energy and water infrastructure; • New scale of AI weaponization.
Social implications (Section 5.4)	• *AI divide*: access to the best-performing agents is expensive; • Widening income gap, particularly in less-industrialized regions; • Enables political manipulation at scale; • Content moderation increases opacity and enables censorship; • Personalized outputs reinforce filter bubbles and polarization.
Cultural consequences (Section 5.5)	• Proliferation of synthetic content; • AI content undermines commons epistemological integrity; • Harder to detect misinformation and propaganda; • Cultural and linguistic homogenization toward WEIRD norms; • Cognitive offloading, particularly among younger users.
Telecommunications (Section 5.6)	• Shift in traffic patterns and network requirements; • 6G sensing introduces pervasive environmental perception; • *Hyper-personalization* masks hyper-surveillance infrastructure; • Expanded cybersecurity attack surface.

6. Limitations and Countervailing Perspectives

Source selection is purposeful rather than systematic: the cases discussed establish that a mechanism operates and is consequential, not how prevalent it is, and the paper presents no primary empirical data. The focus on risks and distributional harms is likewise deliberate, so the account is asymmetric by design and does not weigh benefits against costs.

The benefits of agent intermediation, such as productivity and accessibility for non-expert users, are the reason the trajectory we describe is plausible at all: the layer forms because these systems are useful enough (or at least perceived as such) to be adopted at scale. The configuration we describe is the current one, not the only possible one: the same capabilities can be arranged under different economic and governance conditions, and which arrangement prevails is a matter of political, economic, and social choice rather than of technical necessity. Several countervailing developments fall outside the scope adopted here. AI-mediated access may complement traditional search rather than replace it; retrieval-augmented generation and improving citation practices may restore part of the source attribution we record as reduced [112], although only where the answer is retrieved rather than generated from parametric knowledge, for which no comparable attribution mechanism exists (Section 4.4); and open-weight models, local deployment and open protocols for tool use and inter-agent communication may moderate the concentration described in Section 5.3. These developments qualify our conclusions without displacing the mechanisms we identify, since none alters the position the intermediary occupies between user and source; how far they offset those mechanisms is an empirical question this article cannot settle.

Finally, the causal relationship between agent adoption and the observed traffic declines is empirically suggestive but cannot be fully disentangled from concurrent factors, such as the earlier decline of social media referrals; verifying it would require dedicated empirical work, which future research should undertake.

7. Conclusions

In this paper, we have traced the progression of intermediation of access to online content/services from the technical infrastructure, from the Domain Name System to search engines, social media platforms, super apps, and finally to AI agents. We argue that AI agents constitute a structurally distinct intermediation layer that introduces new dynamics of power, opacity, value distribution, and access to the digital ecosystem, with significant implications for users, online services, and the broader society. The analysis was conducted as a critical, non-systematic review, guided by the boundary questions of Critical Systems Heuristics. The RQs, defined in Section 1, that guided our work ask how AI agent intermediation constitutes a distinct layer from previous forms of internet intermediation and what the implications of this shift are. By answering RQ1, our comparative analysis reveals that AI agent intermediation is structurally distinct from its predecessors: where DNS, search engines, social media platforms and super apps mediated access while leaving users some degree of source visibility and navigational agency, AI agents synthesize responses that can sever the connection to original sources, potentially eliminating both the navigational step and the epistemic signals that previously allowed users to evaluate the provenance and credibility of information, also limiting users' choice. The comparison rests on six structural dimensions and on an explicit criterion distinguishing differences of degree from differences of kind (Table 2); applied across the five layers, it identifies two structural novelties from which the remaining reconfigurations follow: the probabilistic generation of the output and the transfer of enactment to the intermediary. This transformation has profound implications for internet users: the answer to RQ2 reveals the erosion of source transparency, the undermining of open knowledge commons, the concentration of informational power in the hands of a few dominant corporations, impacting not only the economics of content and service provision but also democratic access to information. This includes technology accessibility due to rising costs, but also an infrastructure ready to implement political manipulation at scale and censorship systems. The analysis reveals a consistent distributional asymmetry across all intermediation layers: corporations controlling the intermediation infrastructure capture increasing shares of economic value and user data, while users/citizens, content producers, and open knowledge communities absorb the costs. This asymmetry is most acute in AI agent intermediation, where the content that trained the models generates no traffic or attribution to its creators, threatening the sustainability of the open web infrastructure on which the models themselves depend. The consequences also affect the cultural dimension due to the proliferation of synthetic content, linguistic homogenization, and cognitive offloading, particularly among younger users. Finally, the growing role of task-executing agents as internet intermediaries could affect telecommunications networks in terms of changes to traffic patterns and, consequently, the development of new network management mechanisms.

A particularly productive area of research for both empirical and normative studies is the examination of the governance implications of AI agent intermediation, not addressed in this paper. This involves looking into matters such as agent ownership, attribution mechanisms, compensation models, safety safeguards, and extending platform accountability obligations to AI agent providers.

Author Contributions: Conceptualization, all authors; methodology, M.R. and A.V.; investigation, M.R.; resources, A.V., J.C.D.M., M.B. and G.E.; writing—original draft preparation, M.R.; writing—review and editing, all authors; visualization, M.R.; supervision, A.V., J.C.D.M., M.B. and G.E.; project administration, A.V. and M.B.; funding acquisition, A.V. and J.C.D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by TIM S.p.A.

Data Availability Statement: No new data were created in this study. Data sharing is not applicable to this article.

Conflicts of Interest: M.B. and G.E. are employees of TIM S.p.A., which funded this research under contract no. S25AAFPM 0000205733 (Politecnico di Torino), and contribute to the study as co-authors. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; and in the decision to publish the results. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
DNS	Domain Name System
GenAI	Generative Artificial Intelligence

References

1. De Martin, J.C. Sulla computerizzazione del mondo. *Aut. Aut.* **2025**, *407*, 21–34.
2. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; ukasz Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
3. Cotfas, L.A.; Sandu, A.; Delcea, C.; Diaconu, P.; Frăsineanu, C.; Stănescu, A. From Transformers to ChatGPT: An Analysis of Large Language Models Research. *IEEE Access* **2025**, *13*, 146889–146931. [CrossRef]
4. Bender, E.M.; Gebru, T.; McMillan-Major, A.; Shmitchell, S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 610–623. [CrossRef]
5. Strubell, E.; Ganesh, A.; McCallum, A. Energy and Policy Considerations for Modern Deep Learning Research. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 13693–13696. [CrossRef]
6. Duan, Q.; Lu, Z. AI Agent Communications in the Future Internet—Paving a Path Toward the Agentic Web. *Future Internet* **2026**, *18*, 171. [CrossRef]
7. Wang, L.; Ma, C.; Feng, X.; Zhang, Z.; Yang, H.; Zhang, J.; Chen, Z.; Tang, J.; Chen, X.; Lin, Y.; et al. A Survey on Large Language Model Based Autonomous Agents. *Front. Comput. Sci.* **2024**, *18*, 186345. [CrossRef]
8. Abou Ali, M.; Dornaika, F.; Charafeddine, J. Agentic AI: A Comprehensive Survey of Architectures, Applications, and Future Directions. *Artif. Intell. Rev.* **2025**, *59*, 11. [CrossRef]
9. Bandi, A.; Kongari, B.; Naguru, R.; Pasnoor, S.; Vilipala, S.V. The Rise of Agentic AI: A Review of Definitions, Frameworks, Architectures, Applications, Evaluation Metrics, and Challenges. *Future Internet* **2025**, *17*, 404. [CrossRef]
10. Acemoglu, D.; Restrepo, P. *Artificial Intelligence, Automation and Work*; National Bureau of Economic Research, Inc.: Cambridge, MA, USA, 2018. [CrossRef]

11. Lin, S.; Hilton, J.; Evans, O. TruthfulQA: Measuring How Models Mimic Human Falsehoods. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*; Muresan, S., Nakov, P., Villavicencio, A., Eds.; Association for Computational Linguistics: Dublin, Ireland, 2022; pp. 3214–3252. [CrossRef]
12. Capraro, V.; Lentsch, A.; Acemoglu, D.; Akgun, S.; Akhmedova, A.; Bilancini, E.; Bonnefon, J.F.; Brañas-Garza, P.; Butera, L.; Douglas, K.M.; et al. The Impact of Generative Artificial Intelligence on Socioeconomic Inequalities and Policy Making. *PNAS NEXUS* **2024**, *3*, 191. [CrossRef] [PubMed]
13. Bontridder, N.; Poullet, Y. The Role of Artificial Intelligence in Disinformation. *Data Policy* **2021**, *3*, e32. [CrossRef]
14. Xi, Z.; Chen, W.; Guo, X.; He, W.; Ding, Y.; Hong, B.; Zhang, M.; Wang, J.; Jin, S.; Zhou, E.; et al. The Rise and Potential of Large Language Model Based Agents: A Survey. *Sci. China Inf. Sci.* **2025**, *68*, 121101. [CrossRef]
15. Cambridge Dictionary. Intermediation; Cambridge University Press: Cambridge, UK, 2026. Available online: https://dictionary.cambridge.org/dictionary/english/intermediation (accessed on 27 May 2026).
16. Foster, R. *News Plurality in a Digital World*; Technical Report; Reuters Institute for the Study of Journalism: Oxford, UK, 2012.
17. González-Tosat, C.; Sádaba-Chalezquer, C. Digital Intermediaries: More than New Actors on a Crowded Media Stage. *J. Media* **2021**, *2*, 77–99. [CrossRef]
18. Hutchinson, J. Chapter 2-Digital Intermediation: A Theoretical Framework. In *Digital Intermediation: Unseen Infrastructure for Cultural Production*; Routledge: London, UK, 2023. [CrossRef]
19. Schlosser, M. Agency. In *The Stanford Encyclopedia of Philosophy*; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2015. Available online: https://plato.stanford.edu/entries/agency/ (accessed on 2 October 2025).
20. Lupetti, M.L.; Murray-Rust, D. (Un)Making AI Magic: A Design Taxonomy. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*; Association for Computing Machinery: New York, NY, USA, 2024; pp. 1–21. [CrossRef]
21. Russell, S.J.; Norvig, P. *Artificial Intelligence: A Modern Approach*, 4th ed.; Prentice Hall Series in Artificial Intelligence; Pearson: Boston, MA, USA, 2022.
22. Dorri, A.; Kanhere, S.S.; Jurdak, R. Multi-Agent Systems: A Survey. *IEEE Access* **2018**, *6*, 28573–28593. [CrossRef]
23. Gutowska, A. What Are AI Agents? | IBM. 2024. Available online: https://www.ibm.com/think/topics/ai-agents (accessed on 16 February 2026).
24. Wooldridge, M. Intelligent Agents. In *Multiagent Systems*; Guide Books; MIT Press: Cambridge, MA, USA, 1999; pp. 27–77.
25. Fielding, R.T. Architectural Styles and the Design of Network-Based Software Architectures. Ph.D. Thesis, University of California, Irvine, CA, USA, 2000.
26. Pautasso, C.; Zimmermann, O.; Leymann, F. Restful Web Services vs. "Big"' Web Services: Making the Right Architectural Decision. In *Proceedings of the 17th International Conference on World Wide Web (WWW '08)*; Association for Computing Machinery: New York, NY, USA, 2008; pp. 805–814. [CrossRef]
27. Leitão, P.; Karnouskos, S. *Industrial Agents: Emerging Applications of Software Agents in Industry*; Morgan Kaufmann: San Francisco, CA, USA, 2015.
28. Händler, T. A Taxonomy for Autonomous LLM-Powered Multi-Agent Architectures. In *Proceedings of the 15th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, Rome, Italy, 2023*; SciTePress: Setúbal, Portugal, 2023; pp. 85–98. [CrossRef]
29. Grant, M.J.; Booth, A. A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies. *Health Inf. Libr. J.* **2009**, *26*, 91–108. [CrossRef] [PubMed]
30. Suri, H. Purposeful Sampling in Qualitative Research Synthesis. *Qual. Res. J.* **2011**, *11*, 63–75. [CrossRef]
31. Duboc, L.; McCord, C.; Becker, C.; Ahmed, S.I. Critical Requirements Engineering in Practice. *IEEE Softw.* **2020**, *37*, 17–24. [CrossRef]
32. Ulrich, W. *A Brief Introduction to Critical Systems Heuristics (CSH)*; The Open University: Milton Keynes, UK, 2005.
33. Hutcheson, M.; Morton, A.; Blair, S. Critical Systems Heuristics: A Systematic Review. *Syst. Pract. Action Res.* **2024**, *37*, 499–514. [CrossRef]
34. Berners-Lee, T.; Cailliau, R. WorldWideWeb: Proposal for a HyperText Project. 1990. Available online: https://www.w3.org/Proposal.html (accessed on 27 October 2025).
35. Berners-Lee, T.; Fielding, R.T.; Nielsen, H.F. *Hypertext Transfer Protocol–HTTP/1.0.*; Request for Comments RFC 1945; Internet Engineering Task Force: Wilmington, DE, USA, 1996. [CrossRef]
36. Mockapetris, P. *RFC0882: Domain Names: Concepts and Facilities*; Technical Report RFC1035; Internet Engineering Task Force: Wilmington, DE, USA, 1983.
37. Mockapetris, P.; Dunlap, K.J. Development of the Domain Name System. *SIGCOMM Comput. Commun. Rev.* **1988**, *18*, 123–133. [CrossRef]
38. Boeira, D.F.; Scheid, E.J.; Franco, M.F.; Zembruzki, L.; Granville, L.Z. Traffic Centralization and Digital Sovereignty: An Analysis Under the Lens of DNS Servers. In *Proceedings of the NOMS 2024-2024 IEEE Network Operations and Management Symposium*; IEEE: Piscataway, NJ, USA, 2024. [CrossRef]

39. Bradshaw, S.; DeNardis, L. The Politicization of the Internet's Domain Name System: Implications for Internet Security, Universality, and Freedom. *New Media Soc.* **2018**, *20*, 332–350. [CrossRef]
40. Schiavetta, S.; Komaitis, K. ICANN's Role in Controlling Information on the Internet. *Int. Rev. Law Comput. Technol.* **2003**, *17*, 267–284. [CrossRef]
41. Pearce, P.; Jones, B.; Li, F.; Ensafi, R.; Feamster, N.; Weaver, N.; Paxson, V. Global Measurement of DNS Manipulation. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security 17)*; USENIX: Francisco, CA, USA, 2017; pp. 307–323.
42. Hoang, N.P.; Niaki, A.A.; Dalek, J.; Knockel, J.; Lin, P.; Marczak, B.; Crete-Nishihata, M.; Gill, P.; Polychronakis, M. How Great Is the Great Firewall? Measuring China's DNS Censorship. In *Proceedings of the 30th USENIX Security Symposium (USENIX Security 21)*; USENIX: Francisco, CA, USA, 2021; pp. 3381–3398.
43. Amazon Web Services. Summary of the Amazon DynamoDB Service Disruption in the Northern Virginia (US-EAST-1) Region. 2026. Available online: https://aws.amazon.com/it/message/101925/ (accessed on 26 June 2026).
44. Robinson, D. *A Single DNS Race Condition Brought AWS to Its Knees*; The Register: Londone, UK, 2025.
45. Brin, S.; Page, L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Comput. Netw. ISDN Syst.* **1998**, *30*, 107–117. [CrossRef]
46. Bianchi, T. Global Search Engine Desktop Market Share 2025. Available online: https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/ (accessed on 27 October 2025).
47. Pasquale, F. The Hidden Logics of Search. In *The Black Box Society*; The Secret Algorithms That Control Money and Information; Harvard University Press: Cambridge, MA, USA, 2015; pp. 59–100.
48. Sparrow, B.; Liu, J.; Wegner, D.M. Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. *Science* **2011**, *333*, 776–778. [CrossRef] [PubMed]
49. Fisher, M.; Goddu, M.K.; Keil, F.C. Searching for Explanations: How the Internet Inflates Estimates of Internal Knowledge. *J. Exp. Psychol. Gen.* **2015**, *144*, 674–687. [CrossRef] [PubMed]
50. Evans, D.S. The Economics of the Online Advertising Industry. *Rev. Netw. Econ.* **2008**, *7*, 33. [CrossRef]
51. Edelman, B.; Ostrovsky, M.; Schwarz, M. Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords. *Am. Econ. Rev.* **2007**, *97*, 242–259. [CrossRef]
52. Berman, R.; Katona, Z. The Role of Search Engine Optimization in Search Marketing. *Mark. Sci.* **2013**, *32*, 644–651. [CrossRef]
53. Zuboff, S. *The Age of Surveillance Capitalism*; Profile Books: London, UK, 2019.
54. Lucas, D.; Introna, H.N. Shaping the Web: Why the Politics of Search Engines Matters. *Inf. Soc.* **2000**, *16*, 169–185. [CrossRef]
55. European Commission. *Commission Fines Google €2.95 Billion over Abusive Practices in Online Advertising Technology*; European Commission: Brussels, Belgium, 2025.
56. Plantin, J.C.; Lagoze, C.; Edwards, P.N.; Sandvig, C. Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook. *New Media Soc.* **2018**, *20*, 293–310. [CrossRef]
57. Meta. Meta Reports Fourth Quarter and Full Year 2023 Results. Initiates Quarterly Dividend. Available online: https://investor.atmeta.com/investor-news/press-release-details/2024/Meta-Reports-Fourth-Quarter-and-Full-Year-2023-Results-Initiates-Quarterly-Dividend/default.aspx (accessed on 27 October 2025).
58. Newman, N.; Fletcher, R.; Robertson, C.T.; Ross Arguedas, A.; Nielsen, R.K. *Reuters Institute Digital News Report 2024*; Technical Report; Reuters Institute for the Study of Journalism: Oxford, UK, 2024. [CrossRef]
59. Acemoglu, D.; Ozdaglar, A.; Siderius, J. A Model of Online Misinformation. *Rev. Econ. Stud.* **2024**, *91*, 3117–3150. [CrossRef]
60. Montag, C.; Becker, B.; Gan, C. The Multipurpose Application WeChat: A Review on Recent Research. *Front. Psychol.* **2018**, *9*, 2247. [CrossRef] [PubMed]
61. Bhargava, V.R.; Velasquez, M. Ethics of the Attention Economy: The Problem of Social Media Addiction. *Bus. Ethics Q.* **2021**, *31*, 321–359. [CrossRef]
62. Khalaf, A.M.; Alubied, A.A.; Khalaf, A.M.; Rifaey, A.A. The Impact of Social Media on the Mental Health of Adolescents and Young Adults: A Systematic Review. *Cureus* **2023**, *15*, e42990. [CrossRef] [PubMed]
63. Gorwa, R.; Binns, R.; Katzenbach, C. Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance. *Big Data Soc.* **2020**, *7*, 2053951719897945. [CrossRef]
64. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and Amending Directive 2000/31/EC (Digital Services Act) (Text with EEA Relevance), 2022. *Off. J. Eur. Union* **2022**, *L 277*, 1–102. Available online: https://eur-lex.europa.eu/eli/reg/2022/2065/oj (accessed on 5 November 2025).
65. Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on Contestable and Fair Markets in the Digital Sector and Amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) (Text with EEA Relevance), 2022. *Off. J. Eur. Union* **2022**, *L 265*, 1–66. Available online: https://eur-lex.europa.eu/eli/reg/2022/1925/oj (accessed on 5 November 2025).
66. Mitchell, M. Can Large Language Models Reason? 2023. *AI: A Guide for Thinking Humans*, 10 September 2023. Available online: https://aiguide.substack.com/p/can-large-language-models-reason (accessed on 30 September 2025).

67. Hurler, K. Stack Overflow Traffic Drops as Coders Opt for ChatGPT Help Instead, 2023. *Gizmodo*, 11 May 2023. Available online: https://gizmodo.com/stack-overflow-traffic-drops-as-coders-opt-for-chatgpt-1850427794 (accessed on 2 December 2025).
68. Miller, M. New User Trends on Wikipedia, 2025. *Diff*, Wikimedia Foundation, 17 October 2025. Available online: https://diff.wikimedia.org/2025/10/17/new-user-trends-on-wikipedia/ (accessed on 1 June 2026).
69. Cost, B. Wikipedia Blames ChatGPT for Falling Traffic. *New York Post*, 20 October 2025.
70. Judah, J. How AI and Wikipedia Have Sent Vulnerable Languages into a Doom Spiral, 2025. *MIT Technology Review*, 25 September 2025. Available online: https://www.technologyreview.com/2025/09/25/1124005/ai-wikipedia-vulnerable-languages-doom-spiral/ (accessed on 2 December 2025).
71. Newman, N. *Journalism and Technology Trends and Predictions 2026*; Technical Report; Reuters Institute for the Study of Journalism: Oxford, UK, 2026. [CrossRef]
72. Tang, B.J.; Sun, K.; Curran, N.T.; Schaub, F.; Shin, K.G. Ads That Talk Back: Implications and Perceptions of Injecting Personalized Advertising into LLM Chatbots. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2025**, *9*, 1–47. [CrossRef]
73. Constantino, T. The Studio Ghibli Dilemma–Copyright in the Age of Generative AI, 2025. *Forbes*, 6 May 2025. Available online: https://www.forbes.com/sites/torconstantino/2025/05/06/the-studio-ghibli-dilemma--copyright-in-the-age-of-generative-ai/ (accessed on 8 December 2025).
74. European Union Intellectual Property Office. *The Development of Generative Artificial Intelligence from a Copyright Perspective*; Technical Report; Publications Office of the European Union: Luxembourg, 2025.
75. Miceli, M.; Tubaro, P.; Casilli, A.A.; Bonniec, T.L.; Wagner, C.S.; Sachenbacher, L. *Who Trains the Data for European Artificial Intelligence?* European Parliament-The Lef: Brussels, Belgium, 2024.
76. Karim, M.M.; Khan, S.; Van, D.H.; Liu, X.; Wang, C.; Qu, Q. Transforming Data Annotation with AI Agents: A Review of Architectures, Reasoning, Applications, and Impact. *Future Internet* **2025**, *17*, 353. [CrossRef]
77. Casilli, A.A. *Schiavi Del Clic*; Feltrinelli Editore: Milan, Italy, 2020.
78. Crawford, K.; Joler, V. Anatomy of an AI System: The Amazon Echo as an Anatomical Map of Human Labor, Data and Planetary Resources; AI Now Institute, New York University and Share Lab, Share Foundation: New York, NY, USA, 2018. Available online: http://www.anatomyof.ai (accessed on 25 September 2025).
79. O'Neil, C. AI Can Supercharge Inequality—Unless the Public Learns to Control It. *Nature* **2025**, *646*, 1048–1049. [CrossRef] [PubMed]
80. Weymouth, S. Digital Disintegration: Techno-Blocs and Strategic Sovereignty in the AI Era. *Int. Organ.* **2025**, *79*, S57–S70. [CrossRef]
81. de Vries, A. The Growing Energy Footprint of Artificial Intelligence. *Joule* **2023**, *7*, 2191–2194. [CrossRef]
82. Sajadieh, S.; Fattorini, L.; Perrault, R.; Gil, Y.; Parli, V.; Santarlasci, L.; Pava, J.; Maslej, N.; Altman, R.; Brynjolfsson, E.; et al. *The AI Index 2026 Annual Report*; Technical Report; Institute for Human-Centered AI, Stanford University: Stanford, CA, USA, 2026.
83. Li, P.; Yang, J.; Islam, M.A.; Ren, S. Making AI Less 'Thirsty'. *Commun. ACM* **2025**, *68*, 54–61. [CrossRef]
84. Abdurasulov, A. The New AI Arms Race Changing the War in Ukraine. *BBC*, 10 October 2025.
85. Baker, K.T. AI Got the Blame for the Iran School Bombing. The Truth Is Far More Worrying. *The Guardian*, 26 March 2026.
86. Iraqi, A. 'Lavender': The AI Machine Directing Israel's Bombing Spree in Gaza. *972 Magazines*, 3 April 2024.
87. Zhang, W.; Han, J.; Işık, C.; Song, J.; Liu, X.; Dembińska, I.; Radulescu, M. Inequality in the Digital Economy: The Impact of Artificial Intelligence on Income Gap—An Empirical Analysis Based on County-Level Data from China. *Cities* **2025**, *166*, 106280. [CrossRef]
88. Olejnik, L. AI Propaganda Factories with Language Models, 2025. *arXiv* **2025**, arXiv:2508.20186. [CrossRef]
89. Klincewicz, M.; Alfano, M.; Fard, A.E. Slopaganda: The Interaction between Propaganda and Generative AI. *arXiv* **2025**, arXiv:2503.01560. [CrossRef]
90. Lin, H.; Czarnek, G.; Lewis, B.; White, J.P.; Berinsky, A.J.; Costello, T.; Pennycook, G.; Rand, D.G. Persuading Voters Using Human–Artificial Intelligence Dialogues. *Nature* **2025**, *648*, 394–401. [CrossRef] [PubMed]
91. Diskin, E.I. The use of Artificial Intelligence technologies by internet platforms for the purposes of censorship. *RUDN J. Law* **2024**, *28*, 584–603. [CrossRef]
92. Lazovich, T. Filter Bubbles and Affective Polarization in User-Personalized Large Language Model Outputs. In *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS 2023)*; NeurIPS: San Diego, CA, USA, 2023; pp. 29–37.
93. Saqr, K. Narratives of Divide: The Polarizing Power of Large Language Models in a Turbulent World. *arXiv* **2025.** [CrossRef]
94. Kasy, M. *The Means of Prediction: How AI Really Works (and Who Benefits)*; University of Chicago Press: Chicago, IL, USA, 2025.
95. Chaffer, T.J.; Cotlage, D.; Goldston, J. A Hybrid Marketplace of Ideas. *arXiv* **2025**, arXiv:2501.02132. [CrossRef]
96. Shumailov, I.; Shumaylov, Z.; Zhao, Y.; Papernot, N.; Anderson, R.; Gal, Y. AI Models Collapse When Trained on Recursively Generated Data. *Nature* **2024**, *631*, 755–759. [CrossRef] [PubMed]

97. Brooks, C.; Eggert, S.; Peskoff, D. The Rise of AI-Generated Content in Wikipedia. In *Proceedings of the First Workshop on Advancing Natural Language Processing for Wikipedia*; Lucie-Aimée, L., Fan, A., Gwadabe, T., Johnson, I., Petroni, F., van Strien, D., Eds.; Association for Computational Linguistics: Miami, FL, USA, 2024; pp. 67–79. [CrossRef]
98. Sourati, Z.; Ziabari, A.S.; Dehghani, M. The Homogenizing Effect of Large Language Models on Human Expression and Thought. *arXiv* **2025**, arXiv:2508.01491. [CrossRef]
99. Atari, M.; Xue, M.J.; Park, P.S.; Blasi, D.E.; Henrich, J. Which Humans? *arXiv* **2023.** [CrossRef]
100. Gerlich, M. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies* **2025**, *15*, 6. [CrossRef]
101. Lee, H.P.H.; Sarkar, A.; Tankelevitch, L.; Drosos, I.; Rintel, S.; Banks, R.; Wilson, N. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*; Association for Computing Machinery: New York, NY, USA, 2025; pp. 1–22. [CrossRef]
102. Moon, K.; Green, A.E.; Kushlev, K. Homogenizing Effect of Large Language Models (LLMs) on Creative Diversity: An Empirical Comparison of Human and ChatGPT Writing. *Comput. Hum. Behav. Artif. Hum.* **2025**, *6*, 100207. [CrossRef]
103. Saad, W.; Bennis, M.; Chen, M. A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems. *IEEE Netw.* **2020**, *34*, 134–142. [CrossRef]
104. Stephan, E.; Schott, R.; Lopez, D.; Duan, X.; Morand, L. *AI Agent Protocols for 6G Systems*; Internet Draft draft-stephan-ai-agent-6g-00; Internet Engineering Task Force: Wilmington, DE, USA, 2025.
105. Kunzmann, G.; Khorsandi, B.M.; Ziegler, V.; Raisanen, V.; Ozturk, E.; Kolding, T.; Mannweiler, C. *Technology Innovations for 6G System Architecture*; Technical Report; Nokia Bell Labs: Murray Hill, NJ, USA, 2023.
106. Kirby, J.; Okeleke, A.; Shiko, G.; Stewart, J.; Gabriel, C. *Enabling the Era of Hyper-Personalisation Through Agentic 6G Networks*; Technical Report; Analysis Mason: London, UK, 2025.
107. Karapantelakis, A.; Sahlin, B.; Balakrishnan, B.; Roeland, D.; Rune, G.; Wikström, G.; Wiemann, H.; Arkko, J.; Vandikas, K.; Coldrey, M.; et al. *Co-Creating a Cyber-Physical World*; Technical Report; Ericsson: Stockholm, Sweden, 2024.
108. Pérez, M.G.; Sánchez Sánchez, P.M.; Beltrán, E.T.M.; Vega, F.T.; Karaçay, L.; Paltun, B.G.; Tuna, Ö.F.; Karaman, Ş.; Siniarski, B.; Sandeepa, C.; et al. Strengthening 6G AI/ML Security: Insights from ROBUST-6G Threat Assessment and Prevention Report, 2025. Available online: https://robust-6g.eu/2025/03/05/strengthening-6g-ai-ml-security-insights-from-robust-6g-threat-assessment-and-prevention-report/ (accessed on 10 December 2025).
109. Son, B.D.; Chien, T.V.; Kim, D.I. Trustworthy GenAI over 6G: Integrated Applications and Security Frameworks. *arXiv* **2025**, arXiv:2511.15206. [CrossRef]
110. Manda, J.K. AI-powered Threat Intelligence Platforms in Telecom: Leveraging AI for Real-time Threat Detection and Intelligence Gathering in Telecom Network Security Operations. *Int. J. Multidiscip. Curr. Educ. Res.* **2024**, *6*, 333–340. [CrossRef]
111. Shoetan, P.O.; Amoo, O.O.; Okafor, E.S.; Olorunfemi, O.L. Synthesizing AI'S impact On cybersecurity in telecommunications: A conceptual framework. *Comput. Sci. IT Res. J.* **2024**, *5*, 594–605. [CrossRef]
112. Liu, N.; Zhang, T.; Liang, P. Evaluating Verifiability in Generative Search Engines. In *Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, 2023*; Bouamor, H., Pino, J., Bali, K., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2023; pp. 7001–7025. [CrossRef]