ANALYSIS

Collective Societal Risks: A Blind Spot in AI Governance

Implications for the EU and international trade

April 2026

Executive summary

Artificial Intelligence is a system-transforming technology that is reshaping markets, information flows, and competitive dynamics. The European Union has established a comprehensive regulatory framework that contributes to safe AI, including the AI Act, the Digital Services Act (DSA), and product safety regulation. However, this framework is primarily oriented toward technical risk management, identifiable individual harm, and illegal content.

This analysis argues that AI-driven collective societal risks largely fall between established legal categories. These risks arise through user exposure to AI-enabled products and services and materialise cumulatively through repetition and scale.

The analysis does not propose new legislative instruments. Instead, the ambition is to identify and conceptualise a category of AI-related risks that current EU frameworks are not designed to address. By examining these AI-driven risks from an internal market and trade policy perspective, the study contributes to the broader discussion on how cumulative collective societal risks may influence market order and international trade in a rapidly evolving technological environment.

The analysis distinguishes between three interrelated risk dimensions in which AI-driven collective societal risks manifest:

Informational risks, affecting the integrity of information environments and the distribution of reliable knowledge,

Cognitive risks, relating to long-term effects on how individuals perceive, interpret, and evaluate information,

Behavioural risks, involving systematic influence over users' choices and actions through design and optimisation mechanisms.

Together, these may constitute what the analysis refers to as **cumulative collective societal risks** – risks that do not necessarily qualify as traditional product defects or individual legal harm, but which, due to their scale, may acquire structural significance not only for citizens and consumers, but also for markets, competition, and trade.

This analysis contributes a structured typology of collective societal risks arising from AI-driven behavioural influence mechanisms that aims to translate more abstract societal concerns into the language of existing product and service regulation. By framing informational, cognitive, and behavioural mechanisms as cumulative risks arising from lawful market activity, it highlights why existing EU internal market frameworks struggle to capture their collective societal effects.

From a trade policy perspective, the regulatory blind spot in this area may affect competitive conditions and market access. AI systems' capacity to shape demand, attention, and information exposure implies that competition may increasingly shift from price and quality toward control over behavioural responses and information

environments. This enables firms to tailor pricing and offers based on individual consumer behaviour in ways that were previously much more difficult. In a global AI landscape currently dominated by actors outside the EU, this raises further questions concerning structural dependency, competitive asymmetries, and strategic autonomy.

In digitally mediated markets, AI systems increasingly optimise attention, perception, and behavioural response rather than merely product performance. Empirical research demonstrates that large-scale algorithmic curation and generative AI can shape consumer preferences, information exposure and offline economic behaviour. When deployed globally, these collective societal effects arising from behavioural influence mechanisms may alter competitive conditions, investment incentives, and market access dynamics without changing formal trade rules.

This emerging dimension of AI market power remains largely invisible within EU safety and digital governance frameworks. The analysis concludes that collective societal risks of AI should be explicitly recognised as a relevant risk category within the EU's regulatory frameworks for trade. It further calls for:

- increased awareness-raising among policy makers and regulators about the risks generated by the application of AI technologies on a cumulative, collective level
- a clear operationalisation of behavioural influence mechanisms within the Digital Services Acts' systemic risk regime
- consistent application of the AI Act's provisions concerning manipulative and exploitative practices
- integration of considerations related to risks arising from AI-driven behavioural influence mechanisms into consumer policy, including efforts to strengthen consumer awareness of the risks involved
- strengthened capacity for risk identification, impact assessment, data access, and enforcement
- integration of considerations related to AI-related collective societal risks into international regulatory dialogues and in the WTO and e-commerce discussions.

Table of contents

This analysis is written by Heidi Lund, National Board of Trade Sweden.

Valuable contributions have been made by Per Altenberg, Christian Finnerman, Malin Gisslén, Kristina Olofsson, Isaac Ouro-Nimini Hansen and Katarina Paul at the National Board of Trade as well as Miriam Ben Hadj Ali – Head of Section on AI Policy and Regulation at Ministry of Finance, Docent Didem Gürdür Broo – Associate Senior Lecturer at Department of Information Technology at Uppsala University, and Kristina Knaving – Senior Researcher/Consultant.

1 AI as a transformative technology and emerging societal risks

Artificial Intelligence is here to stay, having already attracted high levels of investment and being increasingly recognized as a transformative technology.[1] The companies developing AI technologies are being valued by markets as the driving forces behind future economic growth, despite the fact that many of the expected economic gains such as enhanced efficiency and higher productivity have yet to materialise in tangible ways. This growing anticipation reflects not merely a technological innovation, but a systemic shift in how economies and markets function, with far-reaching effects for individuals as consumers, workers, and citizens. Understanding the societal implications of this transformation has therefore become an increasingly important policy issue, and is the subject of a growing body of academic and policy analysis.

1.1 Objective and approach

The objective of this analysis is to identify and conceptualise cumulative collective societal risks arising from widespread interaction with AI-enabled goods and services and to explain why existing EU regulatory frameworks struggle to capture these risks. This analysis argues that while the EU has developed an extensive legal framework for AI and digital governance, its reliance on an individual-harm regulatory paradigm leaves collective and cumulative AI-driven societal risks largely unaddressed.

Legal scholarship has increasingly noted that certain AI-related harms emerge at the societal level rather than through identifiable individual injury.[2] Building on this insight, the present analysis examines how such collective societal effects arise through consumer and user interaction with AI-enabled goods and services and how they may influence the functioning of markets, competition, and international trade. The analysis therefore identifies a regulatory blind spot: the absence of an explicit protection objective addressing collective societal risks emerging from large-scale consumer and user interaction with AI systems.

The risks identified in this analysis are also highly relevant in the context of social sustainability and economic security concerns as part of trade policy priorities within the EU[3], but this would require a separate analysis.

---

1 It is possible to argue that an AI boom is real as capabilities are advancing rapidly, applications deliver measurable value, infrastructure is being built and used and long-term transformation is underway. However, there are also concerns of an AI bubble, although views are not unanimous. An AI bubble is motivated by arguments that valuations have disconnected from fundamentals, that circular financing creates illusory demand, that hype far exceeds current capability and that investment is driven by fear of missing out. See analysis of e.g., the European Central Bank and Bank of England.

2 Smuha, N.A. (2021) ‘Beyond the individual: Governing AI’s societal harm’.

3 Trade policy increasingly depends not only on flows of goods and services, but on the social, informational and institutional conditions that sustain open and resilient markets.

1.2 A regulatory blind spot in the EU AI framework

The European Union has adopted a comprehensive legal framework for artificial intelligence and digital services, including the Artificial Intelligence Act (AI Act)[4], the Digital Services Act (DSA)[5] and the General Product Safety Regulation (GPSR).[6] Together with sector-specific product safety legislation and EU cybersecurity law, these instruments establish obligations aimed at promoting trustworthy AI, protecting consumers, and safeguarding fundamental rights within the internal market.[7] However, when examined from the perspective of implementation, market surveillance, and national enforcement, currently high on member states' agendas, a structural shortcoming becomes apparent – the existing regulatory framework is not designed to address collective societal risks arising from AI systems.

Within the current EU framework, AI-driven societal risks are addressed only indirectly and procedurally when they manifest through consumer products or services. Risks that arise as collective societal effects, rather than as identifiable product safety harms or individual rights violations, largely fall outside the regulatory scope.[8] This is not because such risks are unrelated to consumers or markets, but because regulation remains oriented towards individual and measurable harm, while many AI-related risks are collective and cumulative in nature.

Growing concern about societal AI risks is also reflected in member state strategies and policy initiatives that seek to address ethical and societal aspects of AI beyond the current EU regulatory framework. Examples include, but are not limited to, the national AI-strategy of France (AI for Humanity) and the German strategy for AI with the objective of ensuring that AI does not interfere with fundamental human rights and that it is not used in a way that could intensify existing social and economic inequalities.

It is worth acknowledging that certain EU legislative instruments and recent initiatives adopt a more systemic approach to specific categories of societal risk. For example, Regulation (EU) 2024/900 on the transparency and targeting of political advertising recognises that behaviour can be shaped at scale through data-driven advertising systems, including automated optimisation techniques, particularly in the context of elections and democratic debate. However, its scope is limited to political advertising and does not address broader patterns of behavioural influence in digital environments. Similarly, Directive 2005/29/EC concerning unfair business-to-consumer commercial practices in the internal market regulates practices that influence consumer decision-making. Yet it is not framed as a behavioural risk instrument in a broader societal sense; rather, it focuses on economic behaviour, specifically consumer purchasing

---

4 European Union (2024) Regulation (EU) 2024/… of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).

5 European Union (2022) Regulation (EU) 2022/2065 of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act).

6 European Union (2023) Regulation (EU) 2023/988 of the European Parliament and of the Council on general product safety.

7 Commission White Paper on Artificial Intelligence - A European approach to excellence and trust and European Commission (2024) *Europe fit for the Digital Age.*

8 President of the French Republic (2018) *AI for Humanity: French Strategy for Artificial Intelligence*, Federal Government of Germany (2020) *Artificial Intelligence Strategy of the German Federal Government – Update 2020*.

decisions, and not on collective or systemic influence dynamics. Further, emerging initiatives such as the proposed Digital Fairness Act (DFA) signal a welcome and growing EU recognition of behavioural influence in digital environments. However, these efforts primarily aim to strengthen the protection of individual consumer autonomy and fairness in market settings, rather than to regulate cumulative or collective societal behavioural risks.

What remains largely unaddressed are cumulative system-wide effects that arise from the combined impact of many individually lawful platform design and optimisation decisions. While each practice may comply with existing rules, their operation at scale can shape information environments and user behaviour in ways that are difficult to attribute to a single actor or violation. This creates regulatory challenges, as traditional legal frameworks for product and digital safety are designed to address discrete and identifiable harms rather than broader societal influence effects.

It is sometimes argued that regulating design choices in AI systems risks stifling innovation. This analysis does not propose regulating AI design in the abstract. Rather, it highlights a structural regulatory blind spot: where system design choices generate foreseeable population-level effects through scalable optimisation mechanisms, existing frameworks may lack appropriate tools for risk assessment and mitigation.

Comparable regulatory logics exist in other innovation-intensive sectors characterised by systemic risk. Pharmaceuticals, food additives, and financial products are subject to ex ante risk assessment and ongoing safety obligations without preventing innovation; instead, regulatory oversight has shaped innovation trajectories towards safer and more socially sustainable outcomes. From an internal market perspective, there is no clear principled reason to exclude AI-enabled systems whose objective functions demonstrably shape behaviour at scale from similar consideration where population-level effects are foreseeable.

Although collective and cumulative societal risks are difficult to address through regulation, particularly from the perspective of legal responsibility and attribution, there remains a compelling case for broader acknowledgment of their complexity, diversity, and impact.

From a trade-policy perspective, these regulatory blind spots could also translate into structural market disadvantages for EU firms, including in terms of regulatory fragmentation within the EU, market access, and competitive conditions.[9] It is

---

[9] This analysis focuses on identifying and conceptualising cumulative AI-driven societal risks rather than quantifying their trade effects. Establishing a direct empirical relationship between behavioural, cognitive, or informational AI effects and measurable trade outcomes remains difficult, as empirical evidence in this area is still limited. There is research on individual AI-related risk categories, including behavioural influence, as well as on AI and digital trade governance more broadly. Recent studies show that generative AI can produce persuasive misinformation and influence decision-making (Park and Nan, 2025 *Generative AI and misinformation: a scoping review of the role of generative AI in the generation, detection, mitigation, and impact of misinformation*), while trade scholarship highlights how digital interdependence and platform dominance shape trade governance and market access (Kürzdörfer, 2025, *The dog that does not bark – Weaponised interdependence and digital trade at the World Trade Organization* and EPRS, 2025 *The EU's digital trade policy*). However, the interaction between large-scale collective societal risks and the regulation of goods, services, and trade remains underdeveloped in empirical literature. Existing evidence supports the plausibility of such effects, but systematic measurement of trade-level impacts remains limited.

important to recall that one of the purposes of the AI Act was to address fragmentation of national rules, and ensure consistent protection across the internal market.[10]

1.3 Outline and scope of the analysis

The analysis proceeds as follows. **Chapter 1** introduces artificial intelligence as a transformative technology, presents the objective of the analysis, and explains the regulatory challenge posed by collective societal risks arising from widespread interaction with AI-enabled goods and services and their relevance for the internal market and trade policy. **Chapter 2** develops a typology of cumulative collective societal risks arising from behavioural influence mechanisms in AI-enabled goods and services, while **Chapter 3** examines how these risks fall between existing regulatory frameworks. **Chapter 4** examines the structural origins of the regulatory blind spot, analysing the legal and institutional constraints within the EU framework and comparing regulatory approaches to AI-driven influence in the European Union, the United States and China. **Chapter 5** explores the implications of these shortcomings for trade, dependency and strategic leverage in an increasingly geopolitically fragmented AI market, and **Chapter 6** considers the implications for consumer safety and market order. **Chapter 7** concludes with the main findings and policy implications.

As part of a broader trade-policy analysis, this analysis addresses the hidden regulatory risks of AI from the perspective of internal market regulation for goods and services. References to market power, dominance, and global asymmetries are used only descriptively to illustrate trade-relevant effects of the regulatory blind spot, rather than to assess compliance with EU competition law or to propose competition-law remedies. Similarly, references to geopolitics and economic security are made to illustrate the importance of acknowledging the collective societal risks of AI rather than with the aim of demonstrating evidence of the practical impact.

---

[10] The AI Act is adopted on the basis of Article 114 TFEU, which empowers the Union to approximate national laws for the establishment and functioning of the internal market. Recital 3 of the Act clarifies that divergent national AI rules would risk fragmenting the internal market and that uniform Union rules are therefore necessary. The Regulation is thus constitutionally framed as a harmonisation measure aimed at ensuring market coherence while providing a high level of protection.

2 AI, markets, and trade: clarifying the typology of collective societal risks

This analysis argues that the EU's current regulatory architecture systematically underestimates cumulative behavioural, cognitive, and informational risks linked to AI. These are risks that are directly connected to consumer and user exposure but fall between existing legal categories. Here it is important to acknowledge that the regulation of AI is directly interlinked with the regulation of goods and services, therefore also with trade policy.[11] AI-driven behavioural influence mechanisms challenge trade policy because they alter how cross-border markets function, how demand is formed, and how competitive advantages[12] are created and maintained.[13] These effects are particularly relevant in digital services trade, where visibility, ranking, and information exposure determine effective market access.

To understand these dynamics, it is necessary to clarify how AI-related cumulative collective societal risks (hereafter referred to as 'collective societal risks') emerge through consumer interaction with AI-enabled goods and services and how they may accumulate into broader societal and market effects.

2.1 Behavioural influence mechanisms and cumulative collective societal risks of AI

AI-enabled goods and services can influence how individuals perceive information, make decisions, and interact with digital environments. To clarify how these mechanisms operate, it is useful to distinguish between behavioural influence risks affecting individual users and collective societal risks that emerge when such influence operates at scale.

---

[11] The EU Artificial Intelligence Act (AI Act) was notified to the World Trade Organization (WTO) under the Technical Barriers to Trade (TBT) Agreement in 2021 in line with WTO rules requiring WTO members to inform other members about new regulations that could potentially affect trade.

[12] References to market power, dominance, or global asymmetries are used descriptively to illustrate trade-relevant structural effects of regulatory blind spot for goods and services, not to assess compliance with EU competition rules or to propose competition-law remedies.

[13] See, for example: OECD (2023) *The State of Implementation of the OECD AI Principles Four Years On*.

Behavioural influence mechanisms and collective societal risks of AI

Behavioural influence mechanisms (individual level) refer to the ways in which AI-enabled goods and services[14] shape individual cognition, attention, perception, decision-making, and behaviour through design choices such as recommender systems, engagement optimisation, personalisation, behavioural profiling, and asymmetric data accumulation. These mechanisms structure how information is presented, prioritised and amplified, thereby influencing how users interpret and respond to digital environments.

Cumulative collective societal risks (aggregate effects) arise when such behavioural influence mechanisms operate at scale and over time through repeated interaction with AI-enabled goods and services placed on the internal market. These effects are *cumulative* because individually minor, legally permissible, or commercially rational design features may, when cumulative across millions of users and extended over time, gradually amplify and reinforce particular patterns of information exposure, preference formation, and behaviour. Through repeated and interacting algorithmic processes, these systems can produce structural shifts in population-level dynamics. Unlike individual-level harms, these risks are *collective and emergent*: their consequences do not depend on identifiable instances of individual harm, but on cumulative system-level effects mediated by the architecture, optimisation logics and business models of digital platforms operating across the Union.

In this sense, cumulative collective societal risks are not external to consumer protection or market regulation, but emerge as cumulative user risks generated through the large-scale deployment of AI-enabled goods and services within the internal market. Unlike traditional product defects or discrete violations of individual rights, collective societal risks may arise from lawful products and services circulating freely within the internal market. For that reason, they may not be fully captured by existing regulatory risk categories grounded in product safety, individual harm, or case-by-case enforcement logic.

In analytical terms, behavioural influence describes *the mechanism through which effects arise*, while collective societal risks *describe the aggregated outcomes of these effects at scale*. The following section further distinguishes three domains in which these mechanisms manifest: informational, cognitive, and behavioural risks.

---

[14] AI-enabled goods and services encompass *different types of AI systems*, including recommender systems, ranking algorithms and generative AI models such as large language models. These systems operate through different technical mechanisms. *Recommender systems* typically influence behaviour by optimising the ranking and visibility of content or products based on behavioural data, whereas *generative AI systems* primarily influence users through the generation of text, images or other outputs in response to prompts. Unlike recommender systems, generative AI such as large language models, does not personalise outputs to individual users but generate broadly similar responses to user interactions. This results in the societal risk shifting from the fragmentation of shared information environments towards a different concern: the systematic representation of diverse, contested or uncertain knowledge as settled and authoritative, at scale. While the mechanisms differ, both types of systems may contribute to cumulative informational, cognitive, or behavioural societal effects when deployed at scale. It is this scalable influence that gives AI-driven systems potential relevance for market dynamics, competition and trade.

2.2 Typology of AI-driven societal risks

The mechanisms described above can give rise to different but interrelated forms of societal influence.

To clarify these dynamics, it is useful to distinguish between three risk dimensions in which AI-driven collective societal risks manifest. These dimensions reflect different ways in which behavioural influence mechanisms operate and become visible in practice.

Informational risks[15] concern the integrity and structure of information environments, including the large-scale generation and amplification of misleading, biased or false content, as well as asymmetric control over information flows.

Cognitive risks[16] refer to long-term effects on how individuals perceive, process and interpret information, such as epistemic fragmentation, filter bubbles, and the gradual normalisation of distorted or extreme narratives.

Behavioural influence risks[17] relate to systematic influence over user choices and actions, including dependence-inducing design, attention manipulation, nudging and the erosion of autonomous decision-making.

While analytically distinct, these three dimensions of risk reinforce one another and emerge not through isolated instances of harm, but through repeated exposure and widespread use across users and over time.

Together, these may constitute what the analysis describes as **collective societal risks** – risks that do not necessarily qualify as traditional product defects or individual legal harm, but which, due to their scale and societal effects, may acquire structural significance not only for individual citizens and consumers but for markets, competition, and trade.

Perhaps the most significant shortcoming concerns collective and geopolitical dimensions of AI influence. A recurring assumption in EU law is that societal or democratic risks are external to consumer protection and product safety objectives and fall under geopolitical concerns or strategic autonomy. This assumption becomes increasingly difficult to sustain in an AI-driven market.

In regulatory terms, behavioural influence describes the pathway of harm, while societal risk describes its cumulative destination.

---

[15] Example: AI-driven recommender systems on international e-commerce platforms might amplify misleading product reviews or ratings from certain sellers while downranking competitors. This could impact market transparency and the ability of regulators to ensure fair competition across borders.

[16] Example: AI-driven news feeds or social media recommendations may systematically underrepresent success stories of EU tech firms while highlighting negative portrayals. Investors, entrepreneurs, or partners in third countries may develop persistent misperceptions about the EU's technological or economic standing, influencing investment decisions and location choices. This may alter cross-border capital allocation and public support for trade agreements independently of actual fundamentals.

[17] Example: Recommender systems on global digital marketplaces may steer consumers toward a few dominant suppliers via engagement-optimised ranking. Functions as de facto non-tariff measures**,** risk affecting market access, competitive neutrality, and the structure of cross-border digital services trade**.**

AI-related societal risks do not arise as abstract externalities. They are the cumulative result of millions of individual consumer interactions with AI-enabled goods and services placed on the internal market. In regulatory terms, they therefore constitute cumulative societal risks, rather than a separate category of political or geopolitical harm.

EU law already recognises that scale and repetition can transform otherwise limited risks into matters of regulatory concern, as seen in competition law, environmental regulation, and platform governance.[18] AI challenges existing categories, not because it is unrelated to consumer protection, but because it changes the scale, speed, and distribution of risks traditionally associated with products and services.

2.3 Illustrative mechanisms of collective behavioural influence

The societal risks described above become visible through specific mechanisms operating in digital information environments. The following examples illustrate how algorithmic systems can shape information exposure, consumer behaviour, and market dynamics.

[18] Articles 101 and 102 of the Treaty on the Functioning of the European Union (TFEU) regulate anti-competitive agreements and abuse of dominance, respectively. While these provisions do not explicitly reference 'scale' or 'repetition', the application of these rules in practice clearly relies on cumulative market impact – minor practices repeated at scale can distort competition and thus trigger enforcement. Regulation (EU) 2022/2065 (Digital Services Act) introduces graduated rules based on the size and reach of services. The DSA's differentiation between very large online platforms (VLOPs)/very large online search engines (VLOSEs) and other intermediaries is explicitly tied to scale (platforms with >45 million monthly active users). This demonstrates that scale of deployment and repeated transmission of content are regulatory triggers for enhanced obligations (e.g., risk assessments and audits). Further Regulation (EU) 2022/1925 (Digital Markets Act) designates 'gatekeepers' based on quantitative criteria (turnover, user numbers, market capitalisation). Gatekeepers face specific prohibitions and obligations because their scale and repeated market conduct can distort contestability and innovation.

Filter bubbles, epistemic fragmentation, and large-scale reality distortion

Examples of collective societal effects arising from behavioural influence mechanisms[19]

Filter bubbles

A filter bubble arises when individuals are increasingly exposed only to information that aligns with their existing beliefs, preferences, or behaviours, while conflicting or multifaceted perspectives are systematically filtered out often without the user's awareness.

A practical consumer-level example concerns AI-driven recommender systems on social media platforms. So-called 'SkinnyTok' trends illustrate how users who engage with weight-loss or body-image content may quickly be shown increasingly extreme material promoting ultra-thin ideals or restrictive dieting practices. The algorithm interprets even passive engagement such as pausing or rewatching as interest and amplifies similar content, while alternative perspectives (medical advice, recovery content, body-neutral messaging) appear less frequently. The risk does not lie in a single post, but in a self-reinforcing informational environment that shapes perceptions over time. From a consumer-protection perspective, this illustrates how AI systems influence behaviour by structuring exposure raising concerns about informed choice, behavioural autonomy and the integrity of digital information environments.

Similar dynamics appear in *e-commerce*. 'Buy Now, Pay Later' impulse spending loops show how recommender systems learn that urgency-based offers trigger engagement, progressively surfacing high-discount or limited-time products while long-term cost comparisons, quality reviews, or alternative brands appear less prominently. Over time, consumers experience a curated commercial environment that normalises immediate purchasing and short-term financing. In the health and supplement market, search behaviour for remedies can trigger recommender systems to highlight content linking common symptoms to specific brands, detox products, or alternative health claims, while scientific or regulatory perspectives appear less frequently. These examples illustrate how filter bubbles can influence not only beliefs but also concrete consumption patterns.

In economic and trade contexts, algorithmic recommender systems increasingly shape what firms, investors and consumers see. The OECD has documented how data-driven personalisation systems optimise engagement and commercial outcomes by selectively curating content and offers. If AI-driven platforms systematically amplify narratives portraying certain countries as innovation leaders and others as economically stagnant or overregulated, users may be exposed primarily to information reinforcing those frames. For example, if engagement-optimised systems disproportionately surface content suggesting that EU digital regulation suppresses innovation, while downranking analyses highlighting competitiveness gains or consumer protection benefits, audiences may gradually receive a one-sided informational stream. The risk lies not in one article or post, but in a self-reinforcing informational environment that shapes perceptions of trade openness, regulatory quality, and investment attractiveness. Such perception shifts are not merely political; they may influence cross-border investment decisions, consumer trust in foreign providers, and support for trade agreements.

---

[19] The empirical magnitude and societal impact of phenomena such as filter bubbles and epistemic fragmentation remain debated in the academic literature. However, there is broad agreement that large-scale algorithmic curation influences patterns of information exposure and content visibility. The term 'large-scale reality distortion' (also: systemic misinformation) is used here as an analytical concept to describe potential population-level misperceptions arising from large-scale algorithmic amplification; its scale and persistence are subject to ongoing research.

Epistemic fragmentation

Epistemic fragmentation refers to the dissolution of a shared understanding of what constitutes reliable knowledge, which sources are credible, and how truth claims should be evaluated. Rather than a single shared epistemic framework, different groups begin to rely on *distinct and internally coherent* epistemic authorities, norms, and narratives. The problem is therefore not necessarily that information is false, but that *different communities* rely on incompatible sources and standards of validation.

In practice, this may materialise when different professional or political communities rely on different AI-curated sources: one group draws primarily on mainstream economic institutions (OECD, WTO, EU institutions), while another relies on AI-generated summaries derived from partisan outlets or alternative media platforms.[20] Each group trusts different 'experts', accepts different indicators of competitiveness, and interprets trade data through incompatible narratives. Disagreement is no longer only about policy preferences; it concerns the validity of the underlying evidence. Over time, shared deliberation on trade, industrial policy, and regulatory reform becomes more difficult because common epistemic foundations weaken and the shared epistemic foundations necessary for policy deliberation weaken.

Large-scale reality distortion

Large-scale reality distortion on the contrast occurs when AI systems systematically influence how *large segments of the population* perceive facts, events, or social conditions, leading to broadly shared but widespread and persistent misperceptions of reality. There the problem is not fragmentation between groups, but the collective distortion of perceived reality across the information environment. This can occur even without intentional deception
In the EU context, systematically amplified narratives portraying the Union as technologically lagging or economically declining could affect investor confidence, entrepreneurial location decisions and public support for trade agreements. The risk is structural: perceptions influenced by algorithmic amplification may shape voting behaviour, regulatory choices and cross-border capital allocation independently of underlying economic fundamentals.[21] In this respect, AI-driven influence mechanisms may indirectly shape the regulatory and political environment within which trade policy is formulated and market access decisions are made.

These influence mechanisms are not only relevant for individual users or information environments. They also have significant implications for market dynamics, competition, and cross-border trade.

2.4 Market and trade implications of collective societal AI risks

The evidentiary basis for collective societal risks stemming from AI is stronger than is often acknowledged. Large-scale field experiments published in peer-reviewed journals demonstrate measurable effects of algorithmic systems on offline behaviour, emotional states, and political participation.[22] The question is no longer whether these

[20] International organisations such as the World Economic Forum and the European Commission have warned that AI-generated content ecosystems can accelerate informational divergence. See the WEF's 2025 and 2026 Global Risks Reports.

[21] See e.g., Helberger, N., Karppinen, K. and D'Acunto, L. (2018) 'Exposure diversity as a design principle for recommender systems'.

[22] Bond, R.M., Fariss, C.J., Jones, J.J., Kramer, A.D.I., Marlow, C., Settle, J.E. and Fowler, J.H. (2012) 'A 61-million-person experiment in social influence and political mobilization', Nature, 489(7415), pp. 295–298. doi:10.1038/nature11421

effects exist, but whether they are large enough and persistent enough to warrant regulatory attention.

In trade terms, this shifts competition away from price and quality towards control over information, attention, and behavioural response, creating market-shaping de facto non-tariff measures and new forms of market power in digital services trade. For example, by altering market access conditions.[23] AI-driven personalisation and ranking mechanisms weaken core assumptions of perfect competition, notably symmetric information, common price observability and low entry barriers.

Firms can structurally create and exploit information asymmetries through personalised exposure, dynamic pricing and behavioural targeting, while data-driven network effects and learning feedback loops[24] reinforce incumbency advantages and raise structural barriers to entry.[25] Engagement-optimised systems may also generate informational and behavioural externalities that are not internalised in firm-level optimisation, thereby shaping demand formation and competitive outcomes beyond bilateral firm–consumer interactions.[26] In other words in digitally mediated markets, algorithmic visibility functions as a gatekeeping mechanism for competitive opportunities. When control over ranking, recommendation, and data accumulation systematically advantages certain suppliers, the effective conditions of competition may shift independently of formal trade policy measures. This raises trade-policy-relevant questions concerning competitive neutrality and effective market access in cross-border digital services.

In trade terms, algorithmic control over visibility and attention may function as a de facto market access constraint, even in the absence of formal trade barriers. Companies that can harness AI to control information and manipulate attention can give themselves a competitive advantage, effectively locking out others from the market.[27]

AI systems expand firms' ability to implement high-frequency dynamic pricing, personalised offers, and behavioural targeting by processing behavioural and transactional data at scale. Such AI-enabled practices are not just about traditional competition on price and product quality but about leveraging data, behavioural insight and control over information flows to shape consumer preferences and reactions, potentially foreclosing rivals who lack comparable data access or

---

[23] It is also important to acknowledge that issues such as AI's impact on the distribution of power, democracy, inclusion, and human health are related to both social sustainability and economic security as key parameters in international trade.

[24] Feedback loop is a process whereby system outputs influence user behaviour, which generates new data that reinforces the system's future outputs. A feedback loop for generative AI models and their behavioural influence generally operates on a different timescale than recommender systems. Recommender systems operate in short cycles, sometimes real time. Insofar as there is a loop for large language models, the cycles are significantly longer. Engagement data can inform model training, and model output, such as documents, can be used for future training datasets.

[25] Narayanan, A. et al. (2023) 'How Social Media Manipulation Works', Crémer et al. (2019) *Competition Policy for the Digital Era and* Rochet, J.-C. & Tirole, J. (2003) 'Platform Competition in Two-Sided Markets'.

[26] Narayanan et al., 2023; OECD (2025) 'Artificial intelligence and competitive dynamics in downstream markets'.

[27] For example, an AI-driven recommendation engine might prioritise the goods or services of a dominant firm over smaller competitors.

behavioural insights.[28] This enables firms to tailor pricing and offers based on individual consumer behaviour in ways that were previously more difficult. AI-driven control over information flows can also create informational asymmetries, as consumers may be exposed primarily to content that benefits particular products or platforms, thereby distorting competitive conditions.

Collective societal risks from AI and cross-border trade

Algorithmic collective societal risks are not inherently tied to the origin of AI systems. However, cross-border trade expands the scale, speed, and geographic reach of AI-mediated information systems, thereby increasing cumulative societal exposure to algorithmically shaped content. When the most powerful AI systems are developed and controlled outside the EU, this exposure may be structurally amplified.

EU authorities can regulate market access, but they have limited ability to influence the underlying design choices that shape information exposure and behavioural optimisation. This may increase structural dependency, weaken regulatory leverage, and create trade-policy challenges that current EU safety frameworks were not designed to address.[29]

There is further a risk of disproportionate influence on market outcomes: firms with superior access to data, computational resources, and advanced AI systems may shape consumer behaviour at scale, reinforcing incumbency advantages. Smaller firms, particularly those operating from jurisdictions with weaker AI ecosystems or limited access to comparable data and infrastructure, may be unable to replicate these capabilities, potentially creating cross-border competitive imbalances. These risks materialise in the cross-border deployment of AI-enabled goods and services and therefore cannot be addressed solely as individual rights issues or technical safety concerns.[30] Instead, they must be understood as questions of competitive conditions in digital markets, consumer protection, and international trade.[31]

---

[28] OECD (2025) *Artificial Intelligence and Competitive Dynamics in Downstream Markets.* European Parliament Research Service (EPRS) briefing: *Information manipulation in the age of generative artificial intelligence* and European Parliament Research Service (2025) *Information manipulation in the age of generative artificial intelligence.*

[29] Kürzdörfer, N (2025) 'The dog that does not bark – Weaponised interdependence and digital trade at the World Trade Organization' and Aaronson, S.A., (2021) 'Could trade agreements help address the wicked problem of cross-border disinformation?' demonstrating dominance in cross-border digital infrastructures can shape bargaining power and constraints.

[30] EU Commission, White Paper on AI, COM(2020) 65 final.

[31] Addressing a complex regulatory challenge is by no means new. A close parallel can be found in the evolution of chemical regulation in the EU, where early laws focused on specific substances and isolated hazards (much as current EU instruments such as the AI Act, the DSA, and the GPSR address distinct categories of AI-related risk), but later proved insufficient to manage systemic risks arising from widespread and repeated exposure to many chemicals simultaneously. This gap led to a shift towards precautionary, system-oriented frameworks such as REACH, which address aggregate risks across uses and life cycles. Today's fragmented AI regulatory landscape reflects a similar moment, where separate risk-focused regimes may struggle to capture broader societal and behavioural impacts emerging at system level.

3 AI-driven collective societal risks and the limits of existing digital regulation

3.1 The Digital Services Act: strong on content, weak on cognition

The Digital Services Act (DSA) represents a major step forward in platform governance. It addresses illegal content, strengthens transparency obligations, restricts certain forms of advertising (notably towards children), and requires systemic risk assessments for very large online platforms.

However, the DSA primarily regulates what content is distributed, rather than how sustained exposure through algorithmic systems shapes user cognition, behaviour, and perception.[32] Risks such as filter bubbles, epistemic fragmentation, and large-scale distortion of reality are acknowledged as 'systemic risks', but remain insufficiently specified from an enforcement perspective.

Crucially, these risks materialise through ordinary consumer use of digital services. Yet the DSA does not provide enforceable thresholds where cumulative cognitive or democratic erosion, absent illegality triggers regulatory action.

While the DSA introduces important transparency and systemic-risk obligations, it remains primarily focused on content legality and platform process. As a result, cognitive and behavioural effects arising from lawful but pervasive algorithmic exposure remain difficult to address from an enforcement perspective.

3.2 The AI Act: regulating technical risk, not societal power

The AI Act focuses primarily on technical risk management in defined high-risk contexts (e.g., creditworthiness, recruitment, education, and public authority decision-making) and prohibits certain practices under Article 5, including social scoring, specific biometric uses[33] and manipulative or deceptive techniques that materially distort behaviour and cause harm, as outlined inter alia in Article 5.1 (a) and 5.1 (b). These provisions target clearly identifiable unacceptable risks, leaving cumulative behavioural influence arising from lawful large-scale deployment largely outside the Act's scope.

---

[32] The DSA does acknowledge systemic risks to civic discourse, electoral processes, and public security (Art. 34). The weakness lies not in recognition, but in enforceability and justiciability. In the DSA's Explanatory Memorandum**,** it is noted that the act was focused on illegal content and removal procedures but does not directly address the deeper structural issues that algorithms can cause, such as content amplification**,** filtering, or the reinforcement of cognitive biases (which are core to the issues of epistemic fragmentation and distortion of reality). COM/2020/825 final**.** The regulation of algorithmic transparency and its implications for user perception therefore remains limited in addressing these structural dynamics.

[33] GPAI obligations (model cards, training summaries, systemic risk mitigation) exist, but do not address behavioural influence on users.

At the same time, many general-purpose and generative AI systems fall outside the high-risk classification applicable to downstream uses[34] even though they are subject to horizontal obligations. These systems, often not classified as high-risk AI systems under the AI Act nor as products in the traditional product-safety sense, exert influence through widespread interaction with users of AI-enabled goods and services across industries and sectors, including businesses, organisations, and public authorities. These systems may shape information environments through large-scale ranking and recommendation mechanisms, or generate persuasive but potentially misleading content through generative AI models, thereby restructuring information environments across borders.

Societal power, AI, and regulation

In this analysis, **societal power** does not refer to political authority or state-like sovereignty. It refers to the capacity of AI-enabled systems, through large-scale optimisation of information exposure and behavioural responses, to influence patterns of attention, preference formation and decision-making across populations. Unlike traditional product risks, which are typically linked to identifiable defects or individual harm, societal power in this sense emerges cumulatively through repeated consumer and user interactions at scale.

This analysis does not treat societal power as a regulatory object in itself. Rather, it highlights that certain scalable design features, such as engagement optimisation, recommender architectures, behavioural profiling, and asymmetric data accumulation can generate population-level influence effects that may not be fully captured by existing risk categories. Regulatory responses therefore need to focus on identifiable mechanisms and risk pathways rather than abstract notions of power.

While Art. 5.1. (a) and 5.1 (b) of the AI Act may address certain forms of manipulation (e.g., subliminal techniques or exploitation of vulnerabilities), they do not comprehensively regulate all the societal effects that general-purpose AI systems wield, especially when they influence public opinion or societal behaviour on a large scale. This regulatory shortcoming means that large-scale influence effects, which could lead to societal harm, escape both prohibition under Art. 5. 1(a) and 5.1 (b) and high-risk classification under the AI Act.

While Article 50 of the AI Act also introduces transparency requirements for AI systems, ensuring that their capabilities, data usage, and operations are clearly communicated, it does not fully address the societal power these systems exert through large-scale user interaction. Transparency can help identify and monitor the influence of AI systems, such as by clarifying whether content (including images, text, audio, or video) has been manipulated by AI, a requirement under Article 50.4. However, this disclosure does not mitigate the mass dissemination of potentially harmful content. Even if AI-generated content is labelled as such, the ability of AI systems to rapidly spread misinformation, manipulate political discourse, or reinforce

[34] Downstream uses refer to the specific applications or deployment contexts in which an AI system is ultimately used after it has been developed and placed on the market: for example, using a General-Purpose AI model in creditworthiness assessment, deploying an AI system for educational assessment, or embedding a generative model into content moderation or recommender systems.

harmful biases through widespread engagement with users remains largely unregulated. Thus, while transparency is an essential component, it is insufficient in addressing the collective societal effects of AI systems. These systems can still exert influence at a scale that far exceeds individual user-level disclosures, amplifying misinformation or social manipulation in ways that demand a more comprehensive regulatory approach.

These shortcomings are particularly worrying when considering the societal and psychological impacts that may arise from mass engagement with AI systems, including the adoption of AI systems across businesses, organisations and governments that drive policy decisions and market behaviours.

While the AI Act prioritises technical risk management and high-risk use cases, it does not fully account for the societal power exercised by general-purpose and generative AI systems.

A single recommendation is not harmful. A million recommendations, optimised toward the same behavioural outcome, operating continuously across a population, may be. This is precisely the category of risk that falls between existing legal instruments; not because regulators have overlooked it, but because the legal architecture was not designed with cumulative, population-level effects in mind. This oversight creates a significant regulatory blind spot where large-scale influence effects are left unaddressed, raising concerns about the long-term societal consequences of AI systems that escape both prohibition and high-risk classification.

3.3 Behavioural influence risks as a regulatory blind spot

EU product safety law, including the GPSR and sectoral regimes, is built around preventing physical and psychological harm.[35] The AI Act complements this by prohibiting certain forms of direct and manipulative AI practices, such as subliminal techniques that cause significant harm.

Yet a significant category of AI-enabled risks and effects remains insufficiently addressed. These include but are not limited to:

- erosion of user autonomy[36] and decision-making capacity,

---

[35] The General Product Safety Regulation can cover psychological risks, but only where sufficiently concrete, immediate, and attributable. Long-term behavioural influence does not fit enforcement practice. The Regulation defines a 'safe product' as one that, under normal or reasonably foreseeable use, presents 'only the minimum risks compatible with the product's use', and includes mental health effects as part of health and safety considerations. This means that immediate psychological risks (concrete and immediate) that can reasonably be assessed (e.g., causing depression, anxiety, poor sleep) fall within scope of the product's safety risk assessment.

[36] This refers to AI systems or digital platforms subtly shaping or nudging a user's choices so that they are less freely determined, or made under influence or cognitive biases, rather than fully independent reasoning. For example: Recommendation algorithms that prioritise certain content to shape user beliefs (e.g., political, cultural, commercial) or that generative AI that suggests framing, wording, or options that influence decisions without the user realising it.

- dependence-inducing[37]or passivising design,[38]
- gradual normalisation of extreme or distorted viewpoints
- long-term shifts in opinions and behaviour.[39]

Of course, many of these negative effects have existed before AI. However, there is now growing empirical evidence that the widespread deployment of AI systems systematically amplifies these effects at scale.

Generative AI systems may further disrupt information markets[40] by producing persuasive but inaccurate content, reinforcing dominant narratives and dramatically lowering the cost of large-scale manipulation.[41] Documented cases, for example in legal[42] and medical[43] contexts show that degraded information quality can distort financial decisions, investments and consumer choices, with implications for overall market functioning.

Beyond information distortion, AI-driven optimisation of user engagement presents a related but distinct category of risk. AI-driven addictive or engagement-optimised design in apps, games and smart devices is another established risk.[44] Public-health recognition[45] and regulatory guidelines[46] demonstrate that products that formally meet safety standards can nevertheless generate long-term negative behavioural effects, altered consumption patterns, and reduced welfare.

These behavioural effects are not speculative. Large-scale field experiments and academic research show that algorithmic curation and targeted content exposure can

---

37 Design strategies that encourage repeated use, reliance, or habitual engagement**,** potentially reducing proactive decision-making or independent action.

38 The DSA prohibits online platforms from designing or operating interfaces in ways that deceive/ manipulate users or otherwise materially distort or impair their decision-making (Art. 25) and explicitly frames 'dark patterns' as impairing autonomy through interface structure/design/functionalities (Recital 67). For very large platforms and search engines, design-related harms may also be addressed through systemic risk assessment and mitigation duties, including where risks stemming from the design/ functioning of the service lead to serious negative consequences for physical and mental well-being and minors; mitigation may include adapting the design/features/online interface (Arts. 34–35).

39 European Parliamentary Research Service (2025) *Information Manipulation in the Age of Generative Artificial Intelligence*, EPRS Briefing, PE 779.259 EN.

40 I.e., the systems through which information is produced, distributed, priced, discovered, and consumed in the economy, and through which information influences economic decisions.

41 The EPRS briefing, *Artificial intelligence, democracy and elections* describes how general-purpose and generative AI have 'provoked further concerns about the use of artificial intelligence in politics and its impact on the democratic process', noting that synthetic content may comprise most online content by 2026 and raise new risks. See also Stanford Internet Observatory (2024) – *Generative AI and Information Manipulation at Scale*.

42 A recent UK case involved AI hallucinations reaching the High Court, where a barrister's filings were found to contain fabricated citations linked to generative AI use. The court treated it seriously, suggesting that the source was GenAI hallucination and enforcing contempt thresholds.

43 European Heart Journal (2024) – describes a 'use case' where ChatGPT provides realistic but misleading factually incorrect information, illustrating the risk profile in health contexts.

44 See, inter alia, European Commission, DSA proceedings against TikTok concerning the 'TikTok Lite' rewards mechanism (22 April 2024), focusing on systemic risks related to potentially addictive design and effects on minors; and European Commission / CPC Network, consumer-law sweep and coordinated actions on dark patterns, identifying widespread use of manipulative interface design capable of causing behavioural harm despite formal compliance with product safety standards.

45 WHO recognition of 'gaming disorder' in ICD-11.

46 EDPB Guidelines 03/2022 (adopted 14 Feb 2023) on deceptive design patterns in social media platform interfaces. These EU-level guidelines explain how 'deceptive design patterns' steer users and undermine autonomy/consent in social media interfaces (a key institutional reference for 'behaviour-shaping' AI harms).

measurably alter offline behaviour, political participation and emotional states.[47] Related research in digital markets indicates that such behavioural steering mechanisms also influence consumer decision-making and market outcomes, with implications for consumption patterns.[48] Economic research on digital markets further demonstrates that platform design choices reshape content distribution, market structure and welfare outcomes.[49]

EU regulatory assessments under the Digital Services Act support these findings, documenting systemic risks linked to recommender systems, engagement optimisation, and addictive design features. These risks are evidenced through internal company disclosures,[50] academic research,[51] and supervisory investigations,[52] and affect consumer behaviour and digital advertising markets, with potential implications for cross-border digital service markets.

These effects do not arise independently of products or markets. They are the predictable outcome of conscious design choices embedded in AI-enabled goods and services, for example, systems optimised for engagement, attention, or behavioural response. In this sense, behavioural influence is not an externality but an inherent feature of certain business models.

While individual instances of exposure may not meet established thresholds of harm, repetition and scale transform individual risk into collective impact. The economic relevance emerges precisely from this aggregation effect. AI thus challenges the traditional individualised harm model of product safety law, without falling outside its underlying logic.

Finally, asymmetric regulation creates competitive distortions. EU-based firms operate under stringent individual-level consumer and data protection requirements that constrain certain AI design and deployment choices, even though collective societal harms remain largely unregulated. This is not a competition-law assessment of abuse or dominance, but an observation about how differing regulatory constraints translate into structurally different market offerings in international trade.

---

47 There is a substantial and growing body of evidence that algorithmic and AI-enabled systems can systematically influence behaviour at scale, including through controlled experiments on large platforms Bond et al., (2012) 'A 61-Million-Person Experiment in Social Influence and Political Mobilization'; Kramer et al. (2014) 'Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks'.

48 Narayanan et al. (2023) 'How Social Media Manipulation Works'; OECD (2025) *Artificial Intelligence and Competitive Dynamics in Downstream Markets*, Varian (2019) *Artificial Intelligence, Economics, and Industrial Organization*

49 Crémer de Montjoye & Schweitzer (2019) 'Competition Policy for the Digital Era' and OECD (2025) *Artificial Intelligence and Competitive Dynamics in Downstream Markets.*

50 See e.g., The Wall Street Journal – 'The Facebook Files' investigation (2021). This series reports how engagement-driven ranking and platform design can amplify divisive content and shape what users see and do.

51 Milli, Carroll, Wang, Pandey, Zhaob and Dragan (2025) Engagement, user satisfaction, and the amplification of divisive content on social media.

52 Communication from the Commission – Commission Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) C/2024/2537 and Knight-Gilbert Institute & Georgetown University (2025) *Systemic Risk Assessment under the Digital Services Act.*

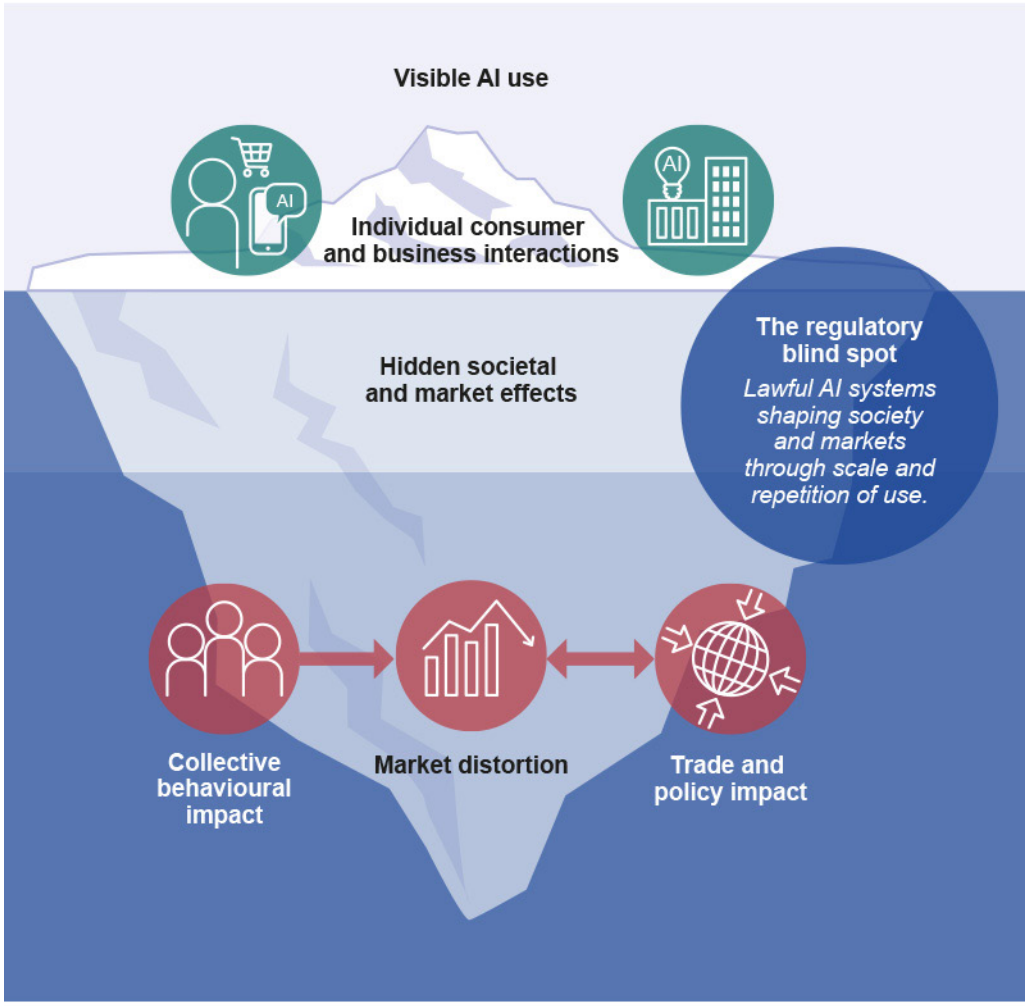
Firms in jurisdictions with weaker consumer protections or stronger state control over information can therefore develop and scale AI systems that EU actors are structurally constrained from offering, and export them into global markets. This regulatory imbalance, increasingly visible in global AI markets, raises questions of fair competition, market access, and AI sovereignty within international trade frameworks. It also highlights that AI's societal impact should not be understood solely or primarily as a free speech issue, but also as a question of market order and economic governance.

As mentioned in Chapter 1, the EU has addressed behavioural influence in the specific contexts (e.g., political advertising[53] and unfair business-to-consumer commercial practices) but does not yet address the structural, collective societal influence effects arising from large-scale AI deployment across markets.

EU product safety and consumer law remain anchored in models of physical or individual psychological harm, leaving collective societal risks arising from behavioural influence mechanisms largely unaddressed.

[53] Regulation (EU) 2024/900, also known as the Regulation on the transparency and targeting of political advertising, is an EU regulation adopted in March 2024.

Figure 1. The regulatory blind spot in AI governance



The existing EU regulatory framework primarily addresses risks related to individual harm and technical safety. However, it does not fully capture lawful AI systems that, through scale and repeated use, shape societal and market outcomes.

These cumulative collective societal effects can be understood along three dimensions:

Effects on consumers and users, including behavioural nudging and dependency, cognitive distortion and reduced autonomy, and increased information asymmetries

Market-level effects, such as shaped demand and preferences, competitive asymmetries, and market lock-in or structural dependency and

Trade and internal market implications, including altered market access conditions, potential structural disadvantages for EU firms, dependency on foreign AI systems, and the emergence of new forms of de facto non-tariff barriers not addressed by traditional trade policy frameworks.

4 Structural origins of the regulatory blind spot

What are the origins of the regulatory blind spot regarding collective societal risks of AI? The structural origins can be explained by the legal and institutional constraints in the EU as well as by an international comparison of approaches in the EU, the US, and China to clarify regulatory differences with respect to AI-driven influence as follows.

4.1 Legal and institutional constraints

The persistence of these grey zones is largely structural.

Legally, behavioural and cognitive harm is difficult to define, prove, and attribute, and raises concerns related to freedom of expression and innovation.

Institutionally, no authority has a clear mandate to protect cognitive autonomy or manage long-term societal impact. Oversight remains fragmented:

- products fall under market surveillance,
- services fall under platform regulation,
- societal risks have no single regulatory owner.

4.2 Comparative approaches: EU, US and China

In order to grasp the starting point for approaches to AI-driven collective influence the following rough comparison can be made between the EU, the US, and China.[54]

[54] The comparison is intended to illustrate how different regulatory models condition the treatment of AI-driven influence in markets and trade, rather than to evaluate their compatibility with competition law or fundamental rights standards.

Table 1. Comparative approaches to AI-driven influence: EU, US and China[55]

Dimension	European Union	United States	China
Primary framing of AI influence	Product and service risks; individual rights; market failures requiring harmonisation	Freedom of speech; national security; geopolitical competition	National security; social stability; ideological control
Regulatory focus	Form, process, and compliance (risk management, transparency, due process)	Strategic power, innovation leadership, and security	Direction, content and societal outcomes
Main regulatory instruments	AI Act; Digital Services Act; product safety and consumer law	State enforcement and civil litigation (including consolidated federal proceedings), alongside competition law and national security or trade instruments.	Algorithm regulation; content controls; direct state oversight
Treatment of societal influence and behavioural control	Addressed only where linked to illegality or individual harm	Constrained by free-speech doctrine; challenged mainly as deceptive/addictive design (esp. for minors)	Explicitly regulated as a societal and security risk
Capacity to address collective societal effects	Limited; collective and cognitive harms weakly operationalised	Limited ex ante; partial ex post leverage via litigation (damages/injunctions), but fragmented and slow	High; intervention at scale is institutionally possible
Ability to intervene against 'lawful but socially harmful' content	Very limited	Minimal; framed as deception/design and youth protection rather than lawful content control	Extensive
Strengths	Legal legitimacy; transparency; proportionality; legal certainty; clear corporate obligations	High freedom of action; rapid technological development; strategic use of AI dominance	Explicit recognition of societal risks; rapid intervention capacity; centralised risk management
Weaknesses	Protects individuals rather than collectives; weak tools for narrative dominance and slow behavioural change	Weak citizen and consumer protection; no clear boundaries against manipulation or disinformation	Absence of individual freedoms; lack of legal due process; institutionalised state manipulation
Underlying trade-off	Legal certainty over influence control	Freedom and power over societal protection	Control and stability over individual rights

[55] See, e.g., Smuha, Nathalie A. & Yeung, Karen (2023) 'The European Union's AI Act', Sloane, Mona & Wüllhorst, Elena (2023) 'A systematic review of regulatory strategies and transparency mandates in AI regulation in Europe, the United States, Canada' and Zou, Mimi & Zhang, Lu (2023) and Zeng, J. (2020) 'Artificial intelligence and China's authoritarian governance – Navigating China's regulatory approach to generative artificial intelligence and large language models'.

The comparison highlights dominant regulatory logics rather than exhaustive legal detail. The comparison highlights a structural dilemma in AI governance. The European Union prioritises legality, procedural safeguards, and individual rights, but lacks effective tools to address collective and behavioural influence. The United States primarily addresses AI-driven influence through a combination of market dynamics, competition enforcement, and national security or trade instruments, while broad consumer-facing regulation of behavioural risks remains limited. However, state consumer-protection enforcement and large-scale civil litigation increasingly function as an ex-post constraint on allegedly addictive or youth-harming design. China, by contrast, explicitly regulates algorithmic influence over information exposure and online behaviour, but does so through centralised state control, inherently incompatible with democratic values and legal due process.

As a result, the EU occupies a distinctive but fragile position: it rejects both laissez-faire influence on markets and authoritarian control, yet has not developed a regulatory model capable of addressing AI-driven societal influence at scale. The EU's regulatory shortcoming is not accidental but structural, rooted in legal constraints, institutional fragmentation, and normative commitments to individual rights. Compared to the US and China, the EU lacks tools to address lawful but socially harmful influence at scale, despite its stronger consumer-protection orientation.

5 AI as a trade dependency: collective societal risks, market power and strategic leverage

It could be argued that the EU, with its comprehensive legal system, is better equipped to address societal challenges than other markets. However, here it must be underlined that most of the global AI ecosystem is currently controlled by companies from the US and China.[56] In practice this implies that EU consumers interact with these platforms from third countries, even if EU regulations exist. It creates a grey zone since while the EU can set rules for firms operating in its jurisdiction, it has only limited ways of controlling the underlying design decisions of foreign AI systems that affect behaviour, cognition, and information consumption.[57]

This chapter provides a discussion of how collective AI-driven societal risks intersect with trade dependency and market power. It examines three interrelated dimensions: limited regulatory control over externally developed AI design, competitive asymmetries arising from large-scale behavioural optimisation, and structural dependency as AI becomes a horizontal input in cross-border value chains. Together, these dynamics illustrate how the regulatory blind spot identified in this analysis may acquire trade-policy relevance.

Regulatory externalisation and limited design control

Beyond formal jurisdictional limits, concentration of AI development outside the Union may also result in *regulatory externalisation*.[58] Design choices, optimisation logics, and technical standards developed in dominant AI ecosystems can effectively structure EU market conditions[59] even where EU law formally applies. In such situations, compliance may consist of marginal adaptation to globally optimised systems, while the underlying architecture shaping information exposure and behavioural optimisation remains externally determined.[60] This dynamic raises trade policy questions concerning regulatory leverage and the Union's capacity to influence the structural parameters of competition in digital markets.[61]

Global scale and competitive asymmetries in AI markets

Collective societal risks such as attention manipulation, misinformation and subtle nudging are amplified when algorithms are globally scaled, making exposure for EU consumers and companies structurally difficult to avoid in integrated digital markets. While the EU's AI Act and GDPR-style frameworks can impose compliance

---

56 Stanford University (2025) *Artificial Intelligence Index Report 2025.*

57 Another way to see it is that the tools might be there within the EU, applying the precautionary principle but at the end of the day, it is a question of balance and proportionality. This implies 'an outsourcing' the task – to the companies themselves to identify and avoid risks which is a dangerous game given that AI as a tech is extremely difficult to control. See also National Board of Trade (2023) *Innovation, AI, Technical Regulation and Trade,* highlighting the challenges regulating digital innovation . A secure application of AI related to behavioural influence is also challenged by the vast technical development and the time to design actual tools for enforcement.

58 OECD (2024) *The Geography of AI Compute: Mapping What Is Available and Where.*

59 European Parliament Research Service (2024) *The Brussels Effect in the Digital Sphere.*

60 Bradford, A., & others (2023) *The EU AI Act's Global Impact and Its Limits.*

61 OECD (2019) *Measuring Distortions in International Markets: The Semiconductor Value Chain.*

obligations and sanctions, regulatory processes operate on a slower timescale than rapid AI innovation cycles. Firms operating under different regulatory approaches may therefore continue behavioural optimisation practices at global scale, potentially generating competitive asymmetries. Where EU-based firms face stringent individual-protection obligations while competing with actors operating under different governance models, concerns regarding competitive neutrality may arise. Such asymmetries do not stem from formal trade barriers, but from divergent approaches to collective societal risks, thereby reinforcing the structural blind spot identified in this analysis.[62]

AI as a horizontal input and structural dependency

When AI becomes a horizontal input resource (like electricity) the effects can materialise in structural dependence in supply chains.[63]

As a general-purpose input embedded in downstream services across sectors including finance, retail, logistics and media, AI-driven optimisation logics may propagate through cross-border value chains. EU firms relying on externally developed AI infrastructure may therefore incorporate behavioural influence mechanisms beyond their direct control, with implications for responsibility allocation, regulatory supervision, and effective market oversight within the internal market.

The EU risks developing a similar structural dependency in regard to AI as it has for semiconductors from a trade-dependency and regulatory-leverage perspective, with high consumption, low control, and a weaker negotiation position. A continued regulatory approach centered on individual harm and firm-level compliance may result in the EU becoming a system user rather than a system owner. In a context of geopolitical tensions, technological dependency in AI may increase the Union's exposure to economic or strategic leverage.[64] To counter this, more forceful means for identification of risks, sanctions, enforcement and consumer awareness raising are avenues that could be considered.[65] However, an introduction of measures must be based on an thorough regulatory impact assessment providing evidence that the effects can be followed-up, measured, and verified not only in relation to increased safety but also in relation to eventual concerns with respect to functioning cross-border trade.

Given the fact that the AI investment landscape is dominated overwhelmingly by funding targeted at AI development, capabilities, and deployment rather than explicit safety work, the question of collective societal effects of AI is critical.[66]

---

62 Crémer, de Montjoye & Schweitzer (2019) 'Competition Policy for the Digital Era'. OECD (2025) *Artificial Intelligence and Competitive Dynamics in Downstream Markets*.

63 Varian, H. (2019) *Artificial Intelligence, Economics and Industrial Organization.*

64 European Commission (2023) *Economic Security Strategy.*

65 For example, if the EU were to rely heavily on AI models or cloud services controlled by non-EU providers for critical sectors such as healthcare, finance, or defence, these providers could condition access, pricing, or service continuity on regulatory or geopolitical considerations.

66 There is formal, expert-oriented documentation indicating that AI safety research funding is an extremely small proportion (ranging from well under 1 %, possibly up to ~2 % with more a generous accounting) of total global AI investment. The most credible quantified estimate comes from the IDAIS policy guide, which explicitly states a < 0.5 % share when safety spending is compared to overall AI R&D budgets. Safe AI Forum (2025) *International Dialogues on Al Safety.* Further, Bengio et al. (2024), in *Science* argue that AI safety research is lagging and that present governance. initiatives lack the mechanisms and institutions to prevent misuse and recklessness, while barely addressing autonomous systems.

6 From individual harm to collective risk: rethinking consumer safety and market order

The current AI and digital governance frameworks are insufficiently equipped to address the most far-reaching AI-related risks to citizens and consumers, not because these risks fall outside the scope of internal market policies, but because they

- do not follow the logic of traditional product defects,
- do not necessarily involve illegal content,
- cannot be reduced to technical malfunctions.

AI-driven societal risks arise through large-scale interaction with AI-enabled goods and services placed on the internal market. Unlike traditional product risks, they do not stem from identifiable defects or illegal conduct, but from the cumulative effects of lawful design choices embedded in digital systems. These design choices[67] shape information exposure and behavioural responses at scale, creating societal risks that existing regulatory frameworks struggle to capture. They concern power, influence, and societal shaping exercised through market relationships.

It should also be acknowledged that various AI-driven collective risks might require different approaches. For information risks such as mis-/disinformation[68] as well as for technical solutions and transparency policy[69] tools might be available. However cognitive and behavioural influence risks[70] and market manipulation via AI-driven actors[71] are much more difficult to manage using standard regulatory tools.[72] Recognising this requires a clearer acknowledgment of how AI transforms the nature of risks associated with products and services in the internal market.

Without such recognition, the EU risks achieving formal regulatory compliance while allowing AI-enabled influence to become an unregulated dimension of competition in the internal market and in international trade. For trade policy, this implies that regulatory blind spots in AI governance can translate into structural market disadvantages that cannot be corrected through traditional trade instruments alone. As a result, serious societal AI risks will escape both product safety enforcement and

---

67 Optimisation logics (engagement, attention, monetisation), not deliberate manipulation.

68 E.g., watermarking, fact-checking, education.

69 For some information and data-related risks, there are established techniques to limit their impact.

70 These are complex because they involve how people psychologically react to AI in everyday decision-making processes. These are systemic, cumulative, and difficult to regulate with traditional legislation.

71 Addressing market manipulation requires extensive international cooperation, monitoring, and new rules in financial markets, not just in AI regulations.

72 For example, filter-bubble dynamics associated with recommender systems do not translate directly to generative AI. Recommender systems optimise explicit engagement metrics (e.g., watch time or click-through rates), which can be audited and potentially regulated because they are clearly specified. Large language models are trained with formal objectives such as helpfulness, but their behaviour largely emerges from training on human preference data rather than explicit engagement metrics. This can produce emergent belief-confirming responses that indirectly drive engagement, but as such dynamics are not tied to explicit engagement targets, they may be harder to detect and regulate.

platform supervision, despite originating in consumer-facing goods and services. A regulatory approach that reacts only to identifiable individual harm risks intervening only after structural and potentially irreversible societal effects have emerged.[73] What can be highlighted, and that has been mentioned initially, is that there are an increasing number of regulatory initiatives that provide a more systemic approach to regulation, integrating behavioural elements which could inspire the regulatory approaches in the area of safe AI for products and services. This could imply making an evaluation of systemic risks mandatory in more areas than today, to better address collective societal risks.

The central policy challenge is therefore not whether AI-driven societal influence exists, but how existing regulatory frameworks for markets, services and trade can recognise and address cumulative collective societal effects that arise from lawful but scalable design features of digital systems.

[73] Compare with the effects of social media-regulation is prepared only when the negative collective consequences are self-evident.

7 Policy conclusions and implications for EU trade and internal market governance

This analysis does not argue that the regulation of AI in the EU is absent – it highlights that the EU's risk framework does not yet capture all relevant societal risks. The central policy challenge is thus the absence of a regulatory vocabulary and governance framework capable of recognising cumulative behavioural influence as a trade-relevant systemic market risk**.** Certain sector-specific EU rules provide building blocks for recognising behavioural influence, but the main regulatory gaps exist in the broader governance of AI-enabled goods and services.

The lack of acknowledgement and regulation of AI-driven collective behavioural influence is not only a theoretical weakness in the EU regulatory framework. It also creates a strategic risk for EU competitiveness, trade policy and strategic autonomy in an environment characterised by trade tensions, technological dominance and a weakened global trade governance framework. Although collective societal risks arising from AI represent a new challenge, recognising these effects and addressing them in policy frameworks and strategic approaches may ultimately contribute to increased trust in goods and services, as well as to the EU's global competitiveness.

Our key policy insights are as follows:

Collective societal risks should be recognised as a form of systemic risk within EU regulatory frameworks governing digital markets and trade-relevant services. EU policy should therefore explicitly acknowledge such risks as a distinct category of risk for consumers and citizens within legislative frameworks governing products and services. AI systems can systematically shape behaviour at scale, generating social and market harms that often fall outside traditional consumer law concepts such as individual detriment, transactional harm, or product defect. Current EU regulation does not fully capture these design-based and collective harms, thereby limiting its ability to address dark patterns, exploitative personalisation, addictive design, or risks affecting vulnerable users, including minors.

Existing frameworks for products and services regulation should be used more systematically and consistently. The DSA's systemic-risk regime should more explicitly operationalise behavioural influence mechanisms (recommender systems, engagement optimisation, addictive design) as core systemic risks, not peripheral issues. These mechanisms should therefore be treated as structural service design features whose foreseeable cumulative effects can generate population-level harm, and should therefore be subject to systematic risk identification, mitigation obligations, and supervisory scrutiny under the DSA.

The AI Act and following guidance should clarify and enforce prohibitions against manipulative and exploitative AI practices as non-compliant, especially those targeting vulnerabilities. Product safety and cybersecurity rules can be applied to AI features where relevant, while recognising their limits for addressing social and psychological harms.

Principles from other regulatory domains can serve as inspiration for strengthening the regulatory approach to safe AI in the products and services domain, particularly with regard to collective societal effects. Given the scale and societal significance of cumulative collective behavioural risk, the effective application of the EU's risk-based regulatory framework will require strengthened research and data-access capacities for risk identification and assessment, as well as sufficient sanctioning and enforcement mechanisms to ensure compliance in practice. Any introduction of new measures must naturally be based on a thorough regulatory impact assessment demonstrating that the effects can be monitored, measured and verified, not only in relation to increased safety, but also with respect to potential impacts on the functioning of cross-border trade.

To make collective societal risks manageable within the EU's risk-based regulatory model, further operational clarification is required. While detailed operational measures are still to be developed, certain principles from existing EU legislation can inform how to address collective behavioural influence in AI governance. For example, Regulation (EU) 2024/900 on the transparency and targeting of political advertising illustrates how transparency and oversight of large-scale, automated influence mechanisms can be structured, while Directive 2005/29/EC concerning unfair business-to-consumer commercial practices in the internal market highlights the importance of protecting vulnerable consumers from undue influence. These precedents suggest that AI regulatory frameworks could include:

- ex-ante assessment of behavioural risks in AI-enabled goods and services,
- obligations for platforms and providers to monitor and mitigate systemic harms,
- mechanisms for supervisory authorities to access relevant data and parameters in a proportionate manner.

The role of consumer information and education in relation to collective behavioural influence mechanisms should be a priority in EU consumer policy, given the difficulty of acknowledging, identifying, and addressing cumulative consumer risks and behavioural influence of AI.

The EU should also promote an international dialogue on AI-driven collective societal risks within the context of international regulatory cooperation as well as within ongoing work in the WTO and e-commerce discussions.

Key concepts and terminology

Addictive design (also: Dependence-Inducing Design)	Design features intended to maximise user retention or repeated interaction.
Algorithmic amplification	The process by which algorithms increase the visibility and reach of specific content based on predicted engagement.
Algorithmic ranking	The ordering of content, products, or services determined by automated systems based on predictive models.
Behavioural targeting	The practice of using behavioural data to tailor content, offers, or advertisements to influence user decisions.
Competitive neutrality (AI context)	The principle that firms competing in a market should operate under comparable regulatory and structural conditions, so that competitive outcomes reflect performance rather than regulatory or systemic advantages.
Data accumulation (also: Asymmetric data accumulation)	The concentration of large-scale behavioural datasets that improve model performance and create *entry barriers* (i.e., structural factors that makes it difficult or costly for new firms to enter and compete effectively in a market).
De facto non-tariff barriers (algorithmic context)	Market constraints that arise not from formal trade restrictions, but from algorithmic systems that influence visibility, ranking, or access to users in ways that may affect competitive opportunities across borders.
Downstream uses	Downstream uses refer to the specific applications or deployment contexts in which an AI system is ultimately used after it has been developed and placed on the market.
Dynamic pricing	Real-time price adjustments based on individual-level behavioural and contextual data.
Engagement optimisation	The process by which an algorithm adjusts outputs to maximise a defined objective function (e.g., engagement, click-through rate, time spent, conversion probability). The objective function is specified during system design and determines which outcomes the system prioritises.
Engagement-optimised systems	Digital systems designed to maximise user engagement metrics, such as time spent, interaction frequency, clicks, or content sharing.
Epistemic fragmentation	The breakdown of shared knowledge structures or common factual reference points within a society, resulting in groups operating with divergent or incompatible understandings of reality. When related to AI systems, it typically concerns algorithmic personalisation, recommender systems, content ranking systems, generative AI content ecosystems or platform-driven information segmentation.
Erosion of user autonomy	This refers to AI systems or digital platforms subtly shaping or nudging a user's choices so that they are less freely determined**,** or made under influence or cognitive biases, rather than fully independent reasoning. For example, recommendation algorithms that prioritise certain content to shape user beliefs (e.g., political, cultural, commercial) or that generative AI that suggests framing, wording, or options that influence decisions without the user realising it.

Feedback loops	A process whereby system outputs influence user behaviour, which generates new data that reinforces the system's future outputs.
Filter bubble	A filter bubble refers to the phenomenon whereby algorithmic personalisation systems selectively expose users to information aligned with their prior preferences, thereby reducing exposure to diverse viewpoints.
General-Purpose AI model (GPAI model)	An AI model that is capable of performing a wide range of different tasks and can be integrated into various downstream systems or applications, typically trained on large and diverse datasets.
Large-scale algorithmic curation	An automated process in which advanced computer programmes (algorithms) select, sort, and present content to individual users on platforms with millions or billions of users, such as Facebook, TikTok, Instagram, and YouTube. This is done to manage enormous volumes of data and deliver a personalised experience that keeps users engaged.
Large scale distortion (also: Systemic misinformation)	In the context of AI systems, large-scale distortion refers to the capacity of algorithmic recommendation, ranking, or generative systems to amplify misleading or inaccurate information at scale, thereby influencing collective understanding. It implies scale and systemic impact, not isolated misinformation.
Market access conditions (digital services)	The practical and regulatory factors that determine whether and how service providers can reach users and compete in digital markets, including platform governance rules, data access, interoperability, and algorithmic ranking mechanisms.
Network effects	A situation where a service becomes more valuable as more users participate.
Nudging	In the AI context, nudging refers to the use of algorithmic design and data-driven personalisation to subtly steer users' behaviour, decisions, or preferences without formally restricting their choice (algorithmically mediated choice architecture). AI-enabled nudging typically operates through recommender systems, interface design, personalised targeting, and engagement optimisation.
Objective function	A mathematical specification of what the AI system is designed to maximise or minimise. For example, engagement, revenue by user, or conversion likelihood (action/transaction probability).
Recommender systems	Algorithmic systems that rank, filter, and prioritise content, products, or information based on predicted user relevance. Algorithmic systems that rank, filter and prioritise content, products or information based on predicted user relevance or engagement, typically using behavioural data and optimisation objectives such as clicks, viewing time or conversion likelihood. These systems influence user behaviour primarily by shaping the visibility and ordering of information.
Structural dependency (AI infrastructure level)	A condition in which markets or jurisdictions rely on external providers for critical AI infrastructure, such as foundational models, cloud computing capacity, or large-scale datasets, thereby limiting strategic autonomy and bargaining leverage.
Two-sided markets	Platforms that serve two interdependent user groups (e.g., consumers and advertisers), where value on one side depends on participation on the other.

References

EU legislation

European Union (2024) Regulation (EU) *2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending certain Union legislative acts (Artificial Intelligence Act)*, *Official Journal of the European Union*, *L 1689, 12 July 2024*.

European Union (2023) *Regulation (EU) 2023/988 of the European Parliament and of the Council of 10 May 2023 on general product safety, amending Regulation (EU) No 1025/2012 and Directive (EU) 2020/1828, and repealing Directive 2001/95/EC and Council Directive 87/357/EEC*, *Official Journal of the European Union*, L 135, 23 May 2023, pp. 1–51. Available at:

European Union (2022) *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)*, *Official Journal of the European Union*, L 277, 27 October 2022, pp. 1–102.

Other references

Aaronson, S.A. (2021). 'Could trade agreements help address the wicked problem of cross-border disinformation?' World Trade Review, 20(3), pp. 420–435.

Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Darrell, T., Harari, Y.-N., Zhang, Y.-Q., Xue, L., Shalev-Shwartz, S., Hadfield, G., Clune, J., Maharaj, T., Hutter, F., Baydin, A.G., McIlraith, S., Gao, Q., Acharya, A., Krueger, D., Dragan, A., Torr, P., Russell, S., Kahneman, D., Brauner, J. and Mindermann, S. (2024) 'Managing extreme AI risks amid rapid progress', *Science*, 381, eadn0117. doi:10.1126/science.adn0117.

Bond, R.M., Fariss, C.J., Jones, J.J., Kramer, A.D.I., Marlow, C., Settle, J.E. and Fowler, J.H. (2012) 'A 61-million-person experiment in social influence and political mobilization', *Nature*, 489(7415), pp. 295–298. doi:10.1038/nature11421.

Bradford, A., et al. (2023) *The EU AI Act's Global Impact and Its Limits* Brookings Institution, Washington DC.

Crémer, J., de Montjoye, Y.-A. and Schweitzer, H. (2019) *Competition Policy for the Digital Era*. Brussels: European Commission, Directorate-General for Competition.

European Central Bank (2025) *Economic Bulletin*, Issue 6/2025. Frankfurt am Main: European Central Bank.

European Commission (2020) *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*. Brussels: European Commission.

European Commission, Joint Research Centre (2021) *The Impact of Online Disinformation on Markets and Economic Behaviour*. Luxembourg: Publications Office of the European Union.

European Commission and High Representative of the Union for Foreign Affairs and Security Policy (2023) *Joint Communication to the European Parliament, the European Council and the Council on the European Economic Security Strategy*. JOIN (2023) 20 final, 20 June 2023. Brussels: European Commission.

European Commission (2024) *Commission Communication – Guidelines for Providers of Very Large Online Platforms and Very Large Online Search Engines on the Mitigation of Systemic Risks for Electoral Processes Pursuant to Article 35(3) of Regulation (EU) 2022/2065*. C/2024/3014, Official Journal of the European Union, 26 April 2024.

European Commission (2024) *Commission Opens Proceedings against TikTok under the Digital Services Act Regarding the Launch of TikTok Lite in France and Spain*. Press release, 22 April 2024.

European Commission (2024) *Digital Services Act: Systemic Risk Assessments and Enforcement Framework*. Brussels: European Commission.

European Commission (2024) *Europe Fit for the Digital Age*. Brussels: European Commission.

European Commission (no date) *Sweeps – Enforcement of Consumer Protection*. Brussels: European Commission.

European Data Protection Board (2023) *Guidelines 03/2022 on Deceptive Design Patterns in Social Media Platform Interfaces*. Brussels: EDPB.

European Parliamentary Research Service (2023) *Artificial Intelligence, Democracy and Elections*. PE 751.478. Brussels: European Parliament.

European Parliament Research Service (2024) *The Brussels Effect in the Digital Sphere* EPRS Briefing. Brussels: European Parliament.

European Parliamentary Research Service (2024) *The Digital Services Act: Systemic Risks and Mitigation*. Brussels: European Parliament.

European Parliamentary Research Service (2025a) *Information Manipulation in the Age of Generative Artificial Intelligence*. PE 779.259 EN. Brussels: European Parliament.

European Parliamentary Research Service (2025b) *The EU's Digital Trade Policy*. Brussels: European Parliament.

Government of Germany (2020) *Artificial Intelligence Strategy of the German Federal Government – Update 2020*. Berlin: Federal Government of Germany.

Helberger, N., Karppinen, K. and D'Acunto, L. (2018) 'Exposure diversity as a design principle for recommender systems', *Information, Communication & Society*, 21(2), pp. 191–207.

Knight-Gilbert Institute and Georgetown University (2025) *Systemic Risk Assessment under the Digital Services Act: Overview*. May 2025.

Kramer, A.D.I., Guillory, J.E. and Hancock, J.T. (2014) ‘Experimental evidence of massive-scale emotional contagion through social networks’, *Proceedings of the National Academy of Sciences*, 111(24), pp. 8788–8790.

Kürzdörfer, N. (2025) ‘The dog that does not bark – Weaponised interdependence and digital trade at the World Trade Organization’, *Review of International Political Economy*, 32(5), pp. 1669–1695.

Legal Cheek (2025) ‘High Court warns lawyers over AI use after “fake” cases cited in submissions’, 9 June.

Milli, S., Carroll, M., Wang, Y., Pandey, S., Zhao, S. and Dragan, A.D. (2025) ‘Engagement, user satisfaction, and the amplification of divisive content on social media’, *PNAS Nexus*, 4(3).

National Board of Trade Sweden (2023) *Innovation, AI, Technical Regulation and Trade: Questioning the Invisible Hand in the Digital Economy*. Stockholm.

Narayanan, A. et al. (2023) ‘How social media manipulation works’, *Journal of Economic Perspectives*, 37(2), pp. 151–174.

OECD (2019) *Measuring Distortions in International Markets: The Semiconductor Value Chain,* OECD Trade Policy Papers No. 234. Paris: OECD Publishing.

OECD (2023) *The State of Implementation of the OECD AI Principles Four Years On*. OECD Artificial Intelligence Papers No. 3. Paris: OECD Publishing.

OECD (2024) *The Geography of AI Compute: Mapping What Is Available and Where.* OECD.AI Policy Observatory, OECD, Paris.

OECD (2025) *Artificial Intelligence and Competitive Dynamics in Downstream Markets*. OECD Competition Roundtables Papers No. 331. Paris: OECD Publishing.

Oxford Internet Institute (2024) *The Effectiveness of Large Language Models for Political Microtargeting*. Oxford: University of Oxford.

Park, S. and Nan, X. (2025) ‘Generative AI and misinformation: A scoping review’, *AI & Society*, 41, pp. 1501–1515.

President of the French Republic (2018) *AI for Humanity: French Strategy for Artificial Intelligence*. Paris.

Rochet, J.-C. and Tirole, J. (2003) ‘Platform competition in two-sided markets’, *Journal of the European Economic Association*, 1(4), pp. 990–1029.

Safe AI Forum (2025) *International Dialogues on AI Safety (IDAIS) Policy Guide*. February 2025.

Siontis, K.C., Attia, Z.I., Asirvatham, S.J. and Friedman, P.A. (2023) ‘ChatGPT hallucinating: Can it get any more humanlike?’, *European Heart Journal*, 45(5), pp. 321–323. doi:10.1093/eurheartj/ehad766.

Sloane, M. and Wüllhorst, E. (2025) ‘A systematic review of regulatory strategies and transparency mandates in AI regulation’, *Data & Policy*, 7, e11. doi:10.1017/dap.2024.54.

Smuha, N.A. (2021) ‘Beyond the individual: Governing AI’s societal harm’, *Internet Policy Review*, 10(3), pp. 1–16.

Smuha, N.A. and Yeung, K. (2023) ‘The European Union’s AI Act’, in *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*. Cambridge: Cambridge University Press.

Stanford Internet Observatory (2024) *Generative AI and Information Manipulation at Scale*. Stanford University.

Stanford University (2025) *Artificial Intelligence Index Report 2025*. Stanford Institute for Human-Centered Artificial Intelligence.

Varian, H.R. (2019) ‘Artificial intelligence, economics and industrial organization’, in Agrawal, A., Gans, J. and Goldfarb, A. (eds.) *The Economics of Artificial Intelligence: An Agenda*. Chicago: University of Chicago Press.

Wall Street Journal (2021) ‘The Facebook Files’. wsj.com/articles/the-facebook-files

World Health Organization (2019) *International Classification of Diseases (ICD-11): Gaming Disorder*. Geneva: WHO.

World Health Organization (2026) *Gaming Disorder: Frequently Asked Questions*. Geneva: WHO.

World Trade Organization (2025) *Harnessing AI for Inclusive Growth*. Geneva: WTO.

World Trade Organization (2026) *WTO Tariff & Trade Data*. Geneva: WTO.

Zeng, J. (2020) ‘Artificial intelligence and China’s authoritarian governance’, *International Affairs*, 96(6), pp. 1441–1459. doi:10.1093/ia/iiaa172.

Zou, M. and Zhang, L. (2025) ‘Navigating China’s regulatory approach to generative artificial intelligence and large language models’, *Cambridge Forum on AI: Law and Governance*, 1, e8. doi:10.1017/cfl.2024.4.

Sammanfattning på svenska
(Summary in Swedish)

AI är en teknik som förändrar marknader, informationsflöden och konkurrensförutsättningar. Europeiska unionen har infört ett omfattande regelverk för säker AI, bland annat AI-förordningen, förordningen om digitala tjänster (Digital Services Act, DSA) och produktsäkerhetslagstiftning. Regelverket är dock främst inriktat på teknisk riskhantering, identifierbar individuell skada och olagligt innehåll.

Denna analys visar att kollektiva samhälleliga risker kopplade till AI i stor utsträckning faller mellan etablerade rättsliga kategorier. Riskerna uppstår genom konsumenters och användares exponering för AI-baserade produkter och tjänster. De utvecklas över tid, när exponeringen sker upprepade gånger och i stor skala.

Syftet med denna analys är inte att föreslå nya rättsliga instrument. I stället är ambitionen att identifiera och begreppsliggöra en kategori av AI-relaterade risker som det nuvarande EU-regelverket för AI inte är utformat för att hantera. Analysen granskar AI-relaterade samhälleliga risker ur ett inre marknads- och handelspolitiskt perspektiv. Den bidrar därmed till en bredare diskussion om hur AI:s kollektiva samhälleliga risker, som utvecklas och förstärks över tid, kan påverka marknadsordningen och den internationella handeln i en snabbt föränderlig teknisk miljö.

Analysen skiljer mellan tre inbördes relaterade riskdimensioner där AI-drivna samhällsrisker manifesteras:

Informativa risker, som påverkar informationsmiljöers integritet och spridningen av tillförlitlig kunskap.

Kognitiva risker, som rör långsiktiga effekter på hur individer uppfattar, tolkar och värderar information.

Beteendemässiga risker, som innebär systematisk påverkan på användares val och handlingar genom design- och optimeringsmekanismer.

Tillsammans kan dessa utgöra det som i analysen benämns **kumulativa kollektiva samhälleliga risker**.

Det handlar om risker som inte nödvändigtvis uppfyller kriterierna för traditionella produktfel eller individuell rättslig skada. Men genom sin omfattning och sina samlade effekter över tid kan de få strukturell betydelse – inte bara för medborgare och konsumenter utan också för marknader, konkurrens och handel.

Analysen bidrar också med en strukturerad typologi av kollektiva AI-drivna samhällsrisker, som översätter mer abstrakta samhälleliga farhågor till språket och perspektivet i traditionell produkt- och tjänstereglering. Genom att betrakta informations-, kognitiva och beteendemässiga mekanismer som kollektiva och kumulativa risker för användare som uppstår till följd av laglig marknadsaktivitet**,** belyser analysen varför befintliga EU-ramverk för säker AI och den inre marknaden har svårt att fånga upp deras samlade och strukturella effekter.

Ur ett handelspolitiskt perspektiv kan en sådan regleringsmässig blind fläck påverka konkurrensvillkor, marknadstillträde, innovationsdynamik och i förlängningen den inre marknadens funktion. AI-system kan forma efterfrågan, styra uppmärksamhet och påverka vilken information användare exponeras för. Det innebär att konkurrensen i allt större utsträckning kan förskjutas från pris och kvalitet till kontroll över beteende-responser och informationsmiljöer.

På digitala marknader optimerar AI-system i allt högre grad uppmärksamhet, perception och beteenderespons snarare än enbart produktprestanda. Empirisk forskning visar att storskalig algoritmisk kuratering och generativ AI kan forma konsumentpreferenser, informationsexponering och ekonomiskt beteende offline. När sådana system används globalt kan dessa aggregerade påverkanseffekter förändra konkurrensvillkor, investeringsincitament och marknadstillträdesdynamik utan att formella handelsregler ändras. Denna framväxande dimension av marknadsmakt förblir i stor utsträckning osynlig inom EU:s regelverk för produktsäkerhet och digital styrning. I ett globalt AI-landskap där många ledande aktörer finns utanför EU väcker detta frågor om strukturellt beroende, konkurrensasymmetrier och strategisk autonomi.

Analysen drar slutsatsen att kollektiva samhälleliga risker till följd av AI:s beteendepåverkan bör erkännas uttryckligen som en relevant riskkategori i EU:s regelverk för handel. Den pekar också på behovet av:

- att skapa större medvetenhet bland beslutsfattare och reglerande myndigheter om riskerna som användningen av AI-teknik medför på kollektiv och aggregerad nivå i samhället
- tydligare kriterier för hur mekanismer för beteendepåverkan ska bedömas inom Förordningen om digitala tjänsters systemiska riskregim
- en konsekvent tillämpning av AI-förordningens bestämmelser om manipulativa och exploaterande praktiker
- att frågor om risker kopplade till AI:s beteendepåverkande mekanismer tydligt integreras i konsumentpolitiken, inklusive insatser för att öka konsumenternas medvetenhet om dessa risker
- stärkt kapacitet för riskidentifiering, dataåtkomst och tillsyn, samt
- att frågor om kollektiva samhälleliga AI-relaterade risker integreras i internationellt regulativt samarbete samt i WTO:s e-handelsdiskussioner.

The National Board of Trade Sweden is the government agency for international trade, the EU internal market and trade policy. Our mission is to facilitate free and open trade with transparent rules as well as free movement in the EU internal market.

Our goal is a well-functioning internal market, an external EU trade policy based on free trade and an open and strong multilateral trading system.

We provide the Swedish Government with analyses, reports and policy recommendations. We also participate in international meetings and negotiations.

The National Board of Trade, via SOLVIT, helps businesses and citizens encountering obstacles to free movement. We also host several networks with business organisations and authorities which aim to facilitate trade.

As an expert agency in trade policy issues, we also provide assistance to developing countries through trade-related development cooperation. One example is Open Trade Gate Sweden, a one-stop information centre assisting exporters from developing countries in their trade with Sweden and the EU.

Our analyses and reports aim to increase the knowledge on the importance of trade for the international economy and for the global sustainable development. Publications issued by the National Board of Trade only reflect the views of the Board.

The National Board of Trade Sweden, April 2026, ISBN: 978-91-89742-85-7

Kommerskollegium
National Board of Trade Sweden

Box 6803, S-113 86 Stockholm, Sweden
Phone +46 8 690 48 00
E-mail registrator@kommerskollegium.se
www.kommerskollegium.se