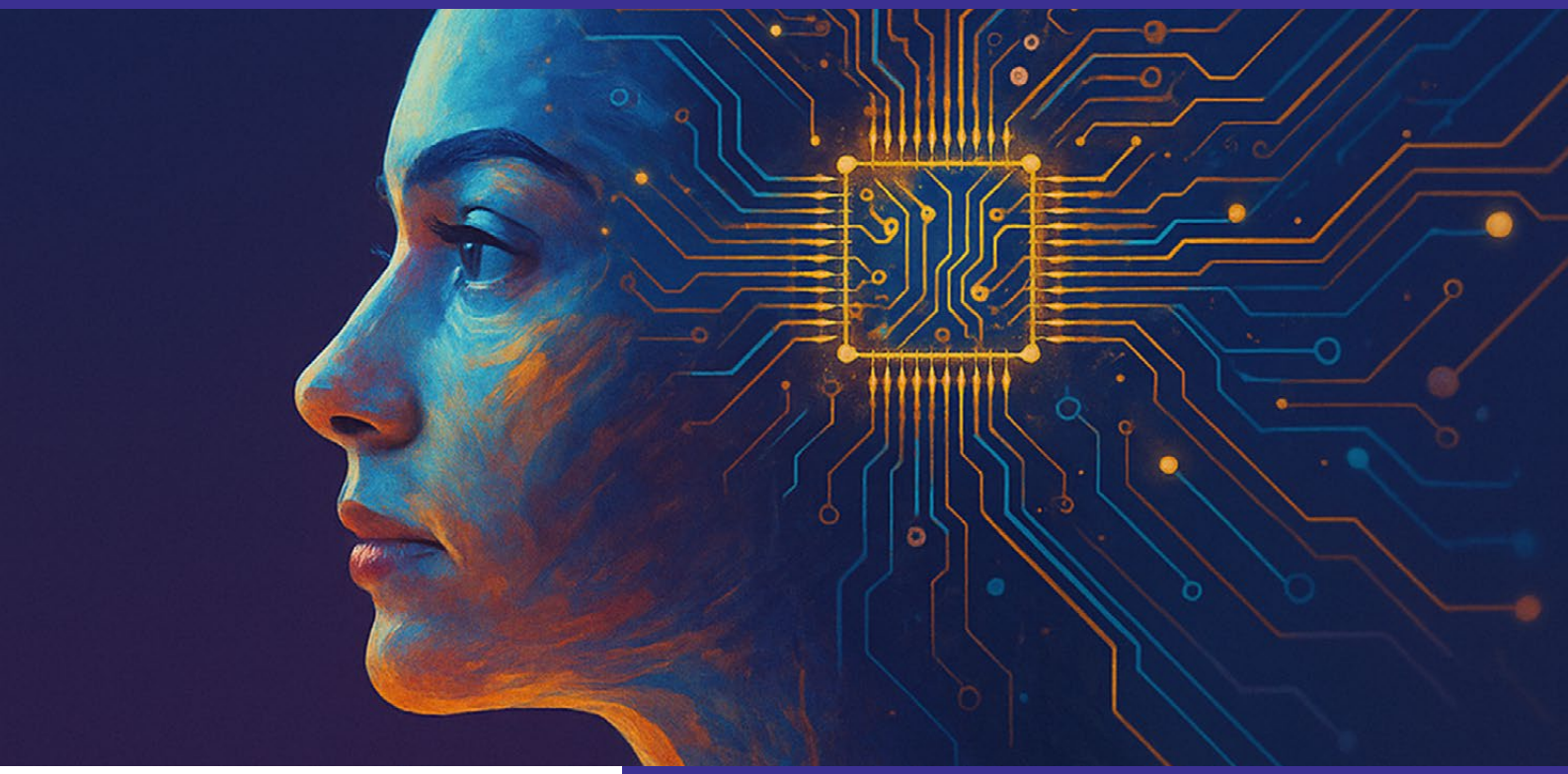
ESCP
BUSINESS SCHOOL

IT ALL STARTS HERE

BERLIN | LONDON | MADRID | PARIS | TURIN | WARSAW



LIGHTS - Technology

Distributed agency in AI-driven organizations: Actor-Network Theory as a lens for responsible integration

ESCP Impact Paper No.2025-36-EN

Sentilkumar GURUMOORTHI & Yannick MEILLER
ESCP Business School

ESCP RESEARCH INSTITUTE OF MANAGEMENT (ERIM)

ESCP Impact Paper No.2025-36-EN

Distributed agency in AI-driven organizations: Actor-Network Theory as a lens for responsible integration[1]

Sentilkumar Gurumoorthi*
Yannick Meiller **

ESCP Business School

Abstract

This paper applies Actor-Network Theory (ANT) to examine how artificial intelligence (AI) systems are integrated into organizations as both actors and socio-technical networks. By focusing on the distributed nature of agency and the opacity introduced by machine learning systems, the paper tackles how responsibility, transparency, and control are reshaped across human and non-human actors. Drawing on examples of AI systems applied to political speech classification and dermatological diagnosis, it shows how blackboxing processes can obscure biases and limit contestability. The paper argues that ANT can guide more ethical and adaptive AI design by supporting critical thinking, stakeholder engagement, and better documentation practices.

Keywords: responsible AI, ANT

* Executive Global PhD student
**Professor, ESCP Business School

1. Chat GPT EDU was used for language purposes.

Distributed agency in AI-driven organizations: Actor-Network Theory as a lens for responsible integration

Introduction

As artificial intelligence (AI) systems become increasingly embedded in organizational practices, questions of agency, accountability, and ethical design grow more pressing. The adoption of machine learning and generative AI tools is transforming not only how decisions are made but also who or what makes them, and how power and responsibility are distributed. Traditional models of control centered on human agency are no longer sufficient to describe these complex sociotechnical dynamics.

This paper proposes to use Actor-Network Theory (ANT) (Latour, 1999) as a conceptual framework for analyzing AI integration within organizations. ANT provides tools to explore how agency is not confined to human decision-makers but is distributed across networks of humans and non-humans: algorithms, data infrastructures, interfaces, and institutional norms. By shifting focus from individual actors to dynamic assemblages, ANT helps uncover how power circulates, how systems become opaque, and how trust is stabilized or undermined.

We argue that ANT not only enhances our analytical grasp of AI systems as both actors and networks but also provides practical insights for more ethical and adaptive implementation. In doing so, we seek to reframe AI not simply as a tool to be deployed but as a sociomaterial phenomenon to be understood, shaped, and governed responsibly.

Actor-Network Theory and distributed agency

Actor-Network Theory (ANT) (Latour, 1999) treats both humans and non-humans (e.g., algorithms, data sets, interfaces, infrastructures...) as "actors" within socio-technical networks. These actors possess the capacity to influence and shape outcomes—not in isolation, but through ongoing associations and negotiations. Rather than assuming a predefined social or technical structure, ANT emphasizes the dynamic formation and reconfiguration of networks through processes of translation, enrollment, and stabilization (Latour, 1987 ; Callon, 1984).

These three processes are central to ANT: *Translation* refers to how actors define roles, align interests, and build alliances to form a network—essentially persuading others to act in ways that support the network's goals. *Enrollment* involves the negotiation and attribution of roles to various actors, both human and non-human, which stabilizes their participation in the network. *Stabilization* occurs when the network becomes robust enough that the roles and relationships are no longer questioned, making the system appear seamless and self-evident.

ANT also distinguishes between *intermediaries* and *mediators* (Latour, 2005). Intermediaries are entities that simply transport meaning or influence without transformation; they are predictable and act as conduits. In contrast, mediators actively transform, translate, or modify the elements they pass along, often producing unexpected outcomes. Recognizing whether an element in the network functions as a mediator or intermediary is crucial for understanding how meaning, agency, and outcomes are constructed and contested within AI systems.

Together, these concepts explain how socio-technical systems gain structure, legitimacy, and momentum within organizations. In the context of AI, this means that decision-making and control are the emergent properties of distributed interactions among users, tools, protocols, and material settings. By shifting attention from individual agents to heterogeneous assemblages, ANT reframes responsibility as relational, collective, and contingent on the alignment of multiple actors across time and space.

AI as actor and network in ANT perspective

Actor-Network Theory allows us to see AI systems not only as actors or nodes within a broader socio-technical network, but also as networks, themselves composite entities composed of many interacting elements. This dual perspective is especially relevant when considering the development and deployment of machine learning systems.

From the ANT viewpoint, an AI application is an actor as it performs roles and influences outcomes, whether by generating recommendations, filtering information, or automating tasks. Yet that same AI system is the product of a networked process involving diverse contributors: data scientists who design the algorithms, engineers who build the infrastructure, managers who define use cases, legal or ethical reviewers who set constraints...

Let's consider training a machine learning system. This involves assembling a temporary but complex network: a specific team, a chosen algorithmic framework, and a training dataset—each with its own origins, assumptions, and actors. The data corpus itself is rarely neutral; it is curated, cleaned, labelled, and interpreted by humans, often reflecting social biases and institutional priorities. These upstream processes shape the capabilities and limitations of the final AI tool, embedding a specific socio-technical configuration within what may appear as a monolithic system.

Now, let's consider using a machine learning system, for example a Generative AI tool (such as ChatGPT, Gemini, Mistral, Claude...). What appears as a simple interface on a user's device is in fact the visible tip of a vast and distributed network. Each interaction mobilizes remote data centers, cloud infrastructure, software environments including trained learning systems maintained by organizations like OpenAI, Google, Mistral AI, etc., and dynamic streams of input and output data. The computation does not happen locally but is the outcome of coordinated activity among distant and hidden actors: hardware, APIs, encryption protocols, monitoring systems, and more. In this sense, every user prompt activates a global assemblage - not a single actor, but rather a networked configuration with deeply embedded power, economic, and technical relations.

By acknowledging AI as both actor and network, ANT makes it possible to uncover the layered construction of intelligence and agency within organizational settings. This perspective deepens our understanding of AI not as an isolated technology but as a crystallization of many interdependent human and non-human contributions, in different places and in different times.

The sources of ethical and responsibility challenges linked to the use of AI are to be found within the folds of this sociotechnical fabric. Identifying relevant structures may shed new light on the issues, to better understand them, spot them in existing systems and find ways to avoid them in future systems. We list a few in the following sections.

Transparency: the central role of machine learning system actors

Transparency is a foundational principle in the pursuit of responsible AI. Transparent AI systems demystify algorithmic processes, helping users understand how decisions are made and why specific recommendations are provided (De Bruijn et al., 2022 ; Mehta et al., 2021). This is especially critical in high-stakes domains such as healthcare, finance, and public services, where opaque systems can erode trust and undermine accountability.

For example, Subramanian et al. (2024) highlight that explainable AI in medical contexts not only enhances collaboration between AI systems and healthcare professionals but also improves decision-making and patient outcomes. From an ANT perspective, transparency is not just about technical explainability but also about making visible the socio-technical network that shapes AI behaviour, including data sources, model assumptions, institutional contexts, and user interfaces.

Achieving meaningful transparency requires identifying the mediators involved in AI decision-making and clarifying their roles. This involves interrogating both human and non-human actors and revealing how their interactions co-produce outcomes. Transparent systems should thus not only disclose internal logic but also support interpretability and contestability across the entire network of actors involved.

Here, a machine learning system plays a major role as its action is irreversible, transforming a learning corpus into a sub-symbolic model—a type of model that does not use explicit symbolic rules but rather encodes knowledge in distributed patterns of parameters. This transformation entails fundamental opacity: users cannot directly access, inspect, or trace the original learning corpus through the final model. The process effectively 'black-boxes' the training data, disrupting transparency between input and output.

In a sociotechnical network, these nodes—machine learning systems—play a central role in shaping (non)transparency. This is all the more true as the team responsible for training the machine learning system is different from the team using it. This suggests that a whole sociotechnical network surrounding this IT node and its dynamics are involved, as developed in the following section.

Transparency: the powerful black-boxing effect of the dynamics of AI as a sociotechnical network

Here we exploit two examples reported in the literature of uses of machine learning which have led to strong biases, widely diffused and hard to denunciate and correct. The ANT analysis shows how the dynamics of a sociotechnical network reinforce and sustain a powerful blackboxing process.

The first case study focused on AI models trained on politically charged speech datasets, particularly speeches by Donald Trump and Kamala Harris (Chalkiadakis et al., 2025). Applying ANT, the human actors identified include data curators, model developers, political analysts, and regulatory agencies, while non-human actors include the speech datasets themselves, classification algorithms, and content moderation policies. Problematization occurred as developers sought to create models capable of distinguishing rhetorical patterns across political ideologies. Interessment strategies included the selective framing of political content and the weighting of language features during model training. Enrollment stabilized when models began categorizing political

speech with high apparent accuracy, leading developers and stakeholders to trust these outputs. However, blackboxing arose as the internal model mechanisms, influenced by dataset biases, became opaque to both users and auditors. Mobilization was hindered because stakeholders outside the original design network (such as civil rights groups) were unable to contest biased classifications meaningfully. The ANT analysis reveals that the network's initial configuration privileged specific linguistic markers, which later resisted scrutiny, complicating transparency and accountability efforts.

The second case study examined dermatological diagnostic AI systems trained predominantly on lighter-skinned patient data (Daneshjou et al., 2022). Here, human actors included dermatologists, AI developers, healthcare providers, and affected patient communities, while non-human actors encompassed image datasets, diagnosis algorithms, and regulatory standards for medical devices. Problematization arose around improving diagnosis accuracy for skin conditions using AI. Interessement involved recruiting datasets largely sourced from existing dermatological image archives, which disproportionately represented lighter skin types. Enrollment was achieved as models demonstrated strong performance metrics on validation sets that mirrored the training data's demographic composition. Blackboxing occurred when these models were deployed clinically; their poor performance on darker skin types was not immediately apparent due to the lack of demographic stratification in testing protocols. Mobilization challenges emerged as marginalized patient groups lacked the mechanisms or data access necessary to contest diagnosis failures. The ANT perspective highlights how non-human actors—specifically, the composition of training datasets—actively shaped the network's stability and the system's perceived transparency, with failures becoming visible only after real-world deployment revealed demographic-specific weaknesses.

In both cases, Actor-Network Theory elucidates how transparency—or its erosion—emerges not from isolated technical flaws but through the configuration and stabilization of heterogeneous actor-networks. Problematization stages defined whose interests were initially prioritized; interessement and enrollment processes reinforced particular data selections and model behaviors; and blackboxing cemented power asymmetries that complicated later efforts at governance and oversight. These insights highlight the need to continuously interrogate human and non-human actor roles throughout the AI lifecycle to foster ethical, transparent, and accountable AI systems.

Critical thinking and AI, or the shift of mediating roles among actors

As AI tools become more deeply embedded in professional and everyday decision-making, the need for critical thinking has grown correspondingly urgent. Too often, users interact with AI systems as if they are neutral tools or infallible oracles, overlooking the complexities of how these systems are built, trained, and maintained. A critical mindset encourages users to question not only the outputs of AI but also the socio-technical networks behind them—the data sources, algorithmic assumptions, institutional biases, and infrastructural dependencies.

Actor-Network Theory reveals the mediating roles AI systems play—not merely transmitting information but actively transforming and reinterpreting it. Critical thinking, in this light, is about recognizing that meaning and authority are negotiated and distributed among many actors, including non-human ones. It also involves understanding that AI tools are situated within power structures and value systems that shape what they prioritize, what they omit, and how they evolve. Cultivating this

perspective is essential not only for more informed use but also for fostering accountability, ethical awareness, and responsible innovation in AI deployment.

To some extent, we have gone from a situation where the human actors could be mediators and the tools they use intermediaries, to a situation in which the AI tools function as mediators and human users may become intermediaries - simply "passing on" the results of the AI system. This deletion - or at least erosion - of the transformative impact of the human actors in the information processing stream is one of the roots of the dangers that come with the increasingly widespread use of AI tools.

Discussion : how ANT can enlighten ethical and design considerations

Transparency and explainability

AI decisions must be explainable—not only to meet regulatory standards but to support effective collaboration. ANT reveals the web of actors involved in producing outputs (data sources, model assumptions, interface cues), encouraging designs that expose these connections. **Ethical design and data security**

Ethical concerns, including questions of bias, fairness, and privacy, are central. ANT reveals how these concerns are embedded in network relations. For instance, biased data is not merely a technical issue but the result of social choices (who collects data, how it's labeled, etc.). ANT-informed design can thus identify leverage points for improving fairness and inclusivity.

Stakeholder engagement

Responsible AI design must include stakeholders—patients, customers, employees—throughout development. ANT supports participatory approaches by mapping how various actors contribute to outcomes. One difficulty here is to think broadly. The case of dermatological diagnosis AI systems shows how the sociotechnical network developing the diagnosis solution failed to include all types of beneficiaries. A sort of open-loop design process prevented them from realising their original datasets were not representative enough of the targeted population. An agile-like approach with regular feedback from real usage, or a broader diversity within the sociotechnical network in charge of developing the solution, would improve the situation.

Importance of metadata

A larger use of metadata could alleviate the unavoidable disruption of transparency between input and output of a machine learning system, and between the learning stage and usage stage of a machine learning system.

These metadata could convey information about the data included in the learning corpus, the priorities which presided over the design stage and the intended scope of use of the final system. In a nutshell, the black box may at least come with a clear manual!

Limitations and future research

This work tackles only the salient structures and dynamics of sociotechnical networks including AI, in order to show the relevance of the ANT framework to understand these systems and help design them.

Now, we need to dig into many cases, in order to identify patterns, either that should be avoided or on the contrary which should be included in future designs.

Also, the ANT analysis highlights the importance of data governance when it comes to AI systems. There is an important avenue of research there.

Conclusion

This paper has shown that Actor-Network Theory (ANT) offers a powerful framework for understanding how AI systems are integrated into organizations, not as isolated technologies but as dynamic configurations of human and non-human actors. AI transforms workflows, decision-making processes, and power dynamics. ANT allows us to move beyond a narrow view of agency and responsibility by highlighting how distributed networks shape action, meaning, and accountability.

We have argued that AI should be viewed both as an actor within larger sociotechnical systems and as a network composed of many actors. This dual perspective is particularly valuable for analysing machine learning systems, whose development and use often involve significant opacity and disconnection between training and operational contexts.

Transparency emerges as a central concern, and ANT helps reveal the conditions under which it is maintained or undermined. Our exploration of two case studies illustrated how powerful blackboxing dynamics can be stabilized through the network's configuration, often resulting in biased and unaccountable outcomes. Critical thinking is therefore essential—not only about outputs but about the systems and assumptions behind them.

ANT also helps reframe ethical and design considerations. It supports more inclusive, reflexive, and participatory approaches to AI development by highlighting whose interests are enrolled and whose are excluded. It stresses the importance of metadata, stakeholder diversity, and documentation practices as means to mitigate the loss of transparency and support systems that are more accountable.

While this paper focuses on conceptual and structural insights, it paves the way for further empirical studies. Mapping out real-world AI networks with ANT can help us identify actionable leverage points for more responsible and adaptive AI integration. Understanding AI through ANT is not merely a theoretical exercise—it is a critical step toward building systems that are not only efficient but also fair, explainable, and accountable.

References

Callon, M. (1984). Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. *The sociological review*, 32(1_suppl).

Chalkiadakis, I., Anglès d'Auriac, L., Peters, G. W., & Frau-Meigs, D. (2025). A text dataset of campaign speeches of the main tickets in the 2020 US presidential election. *Scientific Data*, 12(1), 662

Daneshjou, R., Vodrahalli, K., Novoa, R. A., Jenkins, M., Liang, W., Rotemberg, V., Ko, J., Swetter, S. M., Bailey, E. E., Gevaert, O., Mukherjee, P., Phung, M., Yekrang, K., Fong, B., Sahasrabudhe, R., Allerup, J. A. C., Okata-Karigane, U., Zou, J., & Chiou, A. S. (2022). Disparities in dermatology AI performance on a diverse, curated clinical image set. *Science Advances*, 8(32).

De Bruijn, H., Warnier, M., & Janssen, M. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government* Information *Quarterly*, 39(2), 101666.

Latour, B. (1987). "Science in action: How to follow scientists and engineers through society." Harvard University Press.

Latour, B. (1999). On Recalling ANT. *The Sociological Review*, 47(1_suppl), 15–25.

Latour, B. (2005). "Reassembling the social: An introduction to actor–network-theory." Oxford: Oxford University Press.

Mehta, N., Born, K., & Fine, B. (2021). How artificial intelligence can help us 'Choose Wisely. *Bioelectronic Medicine*, 7(1), 5.

Subramanian, H. V., Canfield, C., & Shank, D. B. (2024). Designing explainable AI to improve human-AI team performance: A medical stakeholder-driven scoping review. *Artificial Intelligence in Medicine*, 149