Neuroethics (2026) 19:33
https://doi.org/10.1007/s12152-026-09661-5

RESEARCH

The Habeas Mentem Principle. Reframing Neurorights in the Digital Age

Gabriele Giacomini

Received: 19 February 2026 / Accepted: 13 July 2026

Abstract This article adopts a liberal perspective on political legitimacy grounded in the protection of freedom and examines the use of digital technologies as instruments of surveillance, censorship, and manipulation. By taking the cases of Egypt (2011) and Iran (2025) as illustrative entry points, the article highlights repressive strategies in the digital age and shows how political power may operate upon individuals' psychological processes. The political dynamics associated with cognitive control and repression reveal the limits of the neurorights paradigm. While necessary to safeguard cognitive liberty and mental privacy, neurorights remain defensive individual rights unable to counter the structural asymmetries generated by states and digital platforms. The paper therefore proposes Habeas Mentem, which incorporates and at the same time transcends neurorights by embedding them within a normative, constitutional, and systemic principle, as well as within a broader framework for balancing cognitive power. This framework requires the separation of digital power, antitrust measures, informational pluralism, and critical education, since mental freedom cannot be secured merely by prohibiting intrusions into the mind, but also by constructing the social and institutional conditions that prevent the emergence of mental domination.

G. Giacomini (✉)
Department of Humanities and Cultural Heritage, University of Udine, Udine, Italy
e-mail: gabriele.giacomini@uniud.it

Keywords Habeas Mentem · Neurorights · Cognitive liberty · Digital surveillance · Informational domination

Introduction

Recent history has demonstrated that digital technologies constitute extremely effective instruments of surveillance, control, manipulation, and repression [1–3]. The evolution of artificial intelligence, automated recognition systems, and, more recently, neurotechnologies extends the power of regimes, often in alliance with private actors, from the physical sphere into the cognitive and mental domain [4–7].

In this scenario, the traditional framework of legal and institutional safeguards risks proving inadequate to protect individual autonomy and citizens' freedom of self-determination [8]. The central problem addressed in this article is therefore the following: how can mental freedom and citizens' decisional autonomy be defended in an era in which political and technological power is capable of intervening in individuals' cognitive, emotional, and decision-making processes? The underlying hypothesis is that liberal democracies must equip themselves with a new normative and institutional principle, capable of complementing the historical achievements of the rule of law and responding to the specific challenges posed by digital technologies and their mental consequences.

Published online: 17 July 2026

The aim of the article is twofold. On the one hand, it seeks to critically analyse the ambivalent role of digital technologies in processes of political mobilisation and authoritarian control, showing how they may weaken, rather than strengthen the right to liberty theorised by the liberal tradition [9, 10]. These risks, although evident in authoritarian regimes, are not entirely absent even in democracies [11, 12]. On the other hand, it proposes the principle of Habeas Mentem as a contemporary extension of classical guarantees of liberty, intended to safeguard mental freedom, cognitive integrity, and citizens' privacy, and to provide a theoretical basis for new rules, rights, and institutions. From this perspective, it becomes necessary to critically interrogate the current paradigm of neurorights [13–15]. They represent an initial attempt to respond to new threats to mental integrity, yet their potential limitation lies in their sectoral nature and in the absence of a systemic politico-social framework.

This article advances the thesis that the protection of mental freedom cannot be ensured through a single regulatory instrument nor through the mere formal recognition of subjective rights. Rather, it requires a multi-level institutional architecture capable of integrating neurorights, algorithmic regulation, antitrust policies, infrastructural transparency, critical education, and the production of cognitive public goods. Only the coordinated action of these mechanisms makes it possible to neutralise the asymmetries generated by the attention economy and the concentration of informational infrastructures, redistributing cognitive power among the state, the market, and civil society. Habeas Mentem therefore assumes a function analogous to that historically performed by Habeas Corpus in past centuries: not merely individual protection against arbitrary interference, but a structural condition of possibility for personal autonomy and democratic self-government [16]. Within this framework, freedom of thought is no longer conceived exclusively as a negative immunity, but as the outcome of a dynamic institutional equilibrium among knowledge, participation, and the mutual control of powers.

Regarding the status of the proposal, Habeas Mentem is presented as an open and interdisciplinary category, capable of being interpreted in partially different ways depending on the perspective adopted. This plurality of formulations is deliberate and reflects the multilayered nature of the problem that Habeas Mentem seeks to address. The forms of cognitive power emerging in digital societies do not belong exclusively to one domain. They concern individual liberty, constitutional guarantees, institutional design, democratic agency, technological infrastructures, and the distribution of epistemic and behavioural influence. For this reason, a concept intended to respond to such transformations cannot be confined in advance to a single disciplinary register, nor organised according to a predetermined hierarchy among its possible interpretations, without losing part of its explanatory and normative potential.

From the standpoint of political philosophy, Habeas Mentem may be understood primarily as a normative ideal aimed at preserving the conditions of mental freedom and democratic autonomy. From a legal standpoint, it may be read as a constitutional meta-principle, capable of guiding the evolution of fundamental rights and the organisation of limits on power in the digital age. From the standpoint of political science, it may be conceived as a systemic framework for analysing and rebalancing the relations among citizens, states, digital platforms, and cognitive infrastructures. At the institutional level, it may also be understood as an architecture of guarantees, responsibilities, and counter-powers designed to protect mental self-determination within technologically mediated environments. Habeas Mentem is, and can legitimately be, all these things. These different formulations are not mutually exclusive, nor do they compete for conceptual priority. Rather, they designate different levels of the same theoretical proposal.

This conceptual openness also performs a methodological function. Since Habeas Mentem addresses problems that cut across political philosophy, constitutional law, political science and technology governance, it allows readers to approach the proposal through the disciplinary lens most relevant to their own specialisation and competences. A legal scholar may therefore emphasise its role as a constitutional principle or meta-principle, a political philosopher may foreground its normative dimension, while a political scientist may find its systemic and institutional implications more salient. This does not make Habeas Mentem conceptually indeterminate. Rather, it defines it as a shared conceptual space within which different disciplinary perspectives can converge, each illuminating a

distinct dimension of the same problem. The polysemy of Habeas Mentem is therefore not an accidental feature of the concept, but part of its theoretical and methodological function. It makes the proposal more adaptable, more interdisciplinary, and better able to capture the complexity of cognitive power in digital societies.

From this perspective, neurorights are situated within a broader architecture of protection and rebalancing of cognitive power. They represent an important component of the Habeas Mentem framework, but they do not exhaust it. While neurorights identify specific legal entitlements connected to mental privacy, cognitive liberty, or protection against manipulation, Habeas Mentem aims to provide the wider normative and institutional horizon within which such rights acquire systemic meaning. This horizon is not limited to rights alone, but also includes institutions, procedures, civic cultures, and mechanisms for balancing economic, social, and political power. Its purpose is not only to defend the individual mind against discrete violations, but also to identify and transform the broader configurations of power that shape the conditions of mental freedom in digital democracies.

The article is structured as follows. The first section reconstructs the framework concerning the governmental use of digital technologies, using suppressed uprisings and protests as illustrative entry points to show that political control does not merely consist of physical coercion or censorship, but may also affect individuals' cognitive processes. The second section introduces the debate on neurorights, reconstructs its theoretical foundations, and highlights its limitations. Although they protect cognitive liberty and mental privacy, they remain necessary but insufficient individual defensive rights in the face of systemic dynamics of informational domination. The third section proposes the paradigm of Habeas Mentem as a constitutional and structural principle for balancing cognitive power, capable of incorporating and surpassing neurorights and articulated through normative principles, separation of technological power, regulation and informational pluralism, together with critical education and cognitive welfare. The conclusion argues that safeguarding mental freedom in the age of artificial intelligence and neurotechnologies requires a comprehensive and systemic institutional transformation.

Egypt, Iran, Cognitive Repression and Manipulation

According to Locke, political power derives its legitimacy from popular consent and has as its primary task the protection of natural rights, among which liberty occupies a central place [17]. When the state fails in this function, thereby losing its legitimacy, citizens are, from a Lockean perspective, fully justified in undertaking forms of resistance [18, 19].

In the contemporary context, digital technologies play a role in the initial phases of popular protest mobilisations, particularly by facilitating coordination, communication, and collective organization [20]. However, they simultaneously provide highly effective instruments to governments, enabling unprecedented modes of surveillance, control, and repression. This dual function of technology highlights a persistent structural tension within contemporary political systems. To illustrate this tension, the section briefly considers the cases of Egypt in 2011 and Iran in 2025. These cases should not be understood as exhaustive or fully comparable case studies, but rather as illustrative entry points. They show what may occur when digital power operates in the absence of robust rights, effective checks and balances, institutional oversight, and a broader equilibrium among political, legal, and technological powers. Under such conditions, digital technologies can become instruments not only of censorship or coercion, but also of cognitive and political domination.

In 2011, protests against the authoritarian regime of Hosni Mubarak erupted in Egypt's major cities, finding their symbolic epicentre in Tahrir Square in Cairo. Thousands of citizens mobilised, demanding social justice and freedom [21]. In this initial phase, the Internet and social networks played a crucial role in facilitating the coordination of protests; at the same time, Egyptian demonstrators soon realised that the network did not represent solely an instrument of emancipation, but also a powerful means of control in the hands of authority [22, 23]. The regime responded with drastic measures, abruptly interrupting nationwide Internet access and silencing millions of users who had until then disseminated dissenting content and established connections abroad [24]. This strategy was subsequently complemented by systems of filtering, selective censorship, and systematic online propaganda activity, aimed at shaping perceptions

and interpretations of events through forms of informational manipulation and the engineering of political emotions.

In this context, a pamphlet began to circulate among activists explicitly recommending the limitation of digital communications, which were considered vulnerable to monitoring by the Ministry of the Interior [25]. As an alternative, the use of analogue tools (such as photocopiers) was suggested for the production and dissemination of leaflets. This strategy reflected a growing awareness of the risks associated with digital exposure in repressive environments, as well as the need to preserve spaces of cognitive autonomy and mental privacy. At the same time, the issue of video surveillance in public spaces and buildings emerged. The pamphlet encouraged demonstrators to neutralise security cameras by rudimentary means, such as spray paint. At the time, artificial intelligence technologies applied to facial recognition and behavioural analysis had not yet reached their current level of sophistication. Today, such systems enable continuous, automated, and predictive monitoring based on machine-learning models and deep neural networks, thereby rendering surveillance more pervasive.

The second case concerns the mobilisations in Iran, which erupted at the end of December 2025 and form part of a long tradition of cyclical waves of protest and popular mobilization [26, 27]. The protests began in Tehran's Grand Bazaar and rapidly spread to more than 110 cities, involving broad segments of the population. Although initially triggered by economic causes, the mobilisations soon developed into a wider political movement, characterised by criticism of religious leadership and demands for structural reforms. The horizontal movement also received symbolic support from figures in exile, including Reza Pahlavi [28]. In response, the Iranian government adopted a combination of repressive measures and containment tactics: systematic use of force, mass arrests, and attempts to control the public narrative [29].

The arena for political action progressively expanded from the streets to the digital sphere. For the insurgents, tools such as encrypted messaging applications, VPN networks, and, where possible, satellite connections constituted useful channels for communication, coordination, and the documentation of abuses by security forces. However, the same technologies were also used by the state to monitor, track, and repress dissent. Iranian authorities imposed near-total communication blackouts in order to isolate the population. Such interruptions not only conceal human rights violations but also constitute, in themselves, a serious infringement of the right to information and online freedom of expression, with direct effects on individuals' capacity for informational orientation.

In parallel, the state intensified its use of advanced digital surveillance systems, including facial recognition, behavioural analysis, interception, and forced access to personal devices, in order to identify and neutralise activists and dissidents [30]. Prolonged exposure to contexts of pervasive surveillance has the deliberate objective of fostering chronic states of hypervigilance, fear, anxiety, and internalised self-censorship. Alongside these instruments, there is a systematic use of digital propaganda and targeted disinformation. Through networks of coordinated accounts, bots, and controlled media channels, the regime promoted narratives aimed at delegitimising protesters, portraying them as agents of foreign powers or as destabilising elements. The objective of such practices is to selectively shape the population's perceptions and beliefs [31].

The cases of Egypt and Iran do not constitute isolated episodes, but rather form part of a broader global pattern that also includes recent mobilisations in countries such as Myanmar, Hong Kong, and Belarus [8]. Authoritarian regimes systematically exploit digital technologies for purposes of surveillance, propaganda, censorship, and repression of fundamental freedoms. Such practices may assume a predominantly reactive dimension, oriented towards the direct containment of dissent, or a proactive dimension, aimed at the cognitive shaping of citizens' conduct and opinions.

A paradigmatic example of a reactive strategy is represented by the blocking of digital networks under particular conditions (for instance, during elections or demonstrations) or by the deployment of surveillance systems based on artificial intelligence, especially facial recognition networks [32–34]. The application of machine-learning algorithms to neural data makes it possible, at least in theory, to infer political preferences, ideological orientations, and latent intentions. By contrast, proactive strategies are grounded in the systematic use of digital propaganda, supported by psychographic techniques and microtargeting made

possible through the mass collection of personal data and algorithmic processing [35]. Messages are constructed in a personalised manner on the basis of psychological profiles, cognitive habits, and emotional vulnerabilities, in order to maximise persuasive effectiveness and to enhance the stability and reputation of the political regime [36]. In future perspective, the development of neurotechnologies may further reinforce these dynamics. Brain–computer interfaces, both invasive and non-invasive, based on the analysis of neural signals, could enable a more direct influence on decision-making processes, raising fundamental ethical questions concerning the right to mental freedom and cognitive self-determination.

The transition from digital repression to cognitive repression should be understood in a specific sense. Censorship, surveillance, internet shutdowns, and algorithmic propaganda do not affect only external communication, but also the conditions through which individuals orient themselves, form judgements, regulate fear, and decide whether to act politically. Shutdowns produce isolation and informational disorientation; surveillance fosters fear, hypervigilance, and self-censorship; propaganda and disinformation shape perceptions and emotions; and microtargeting acts upon attention and individual vulnerabilities. Although citizens are not passive and there is not always a direct and deterministic relationship between repression and its effects, digital repression generally alters the cognitive and affective conditions of political autonomy, and is often implemented precisely for this purpose.

The decision to introduce the problem by starting from certain authoritarian contexts serves a specific diagnostic function. These contexts are not taken as the principal reference point for Habeas Mentem, but rather as limiting cases capable of making visible, in an extreme form, the risks towards which liberal democracies may also drift when digital technologies operate in the absence of effective rights, independent oversight, informational pluralism, and adequate institutional checks and balances. Habeas Mentem is therefore primarily relevant to democracies, before initially fragmented, exceptional, or sectoral forms of cognitive power consolidate into ordinary structures of domination. Fortunately, liberal democracies do not coincide with the Egyptian or Iranian scenarios, but they may nevertheless experience gradual, opaque, and ambiguous processes of concentration of cognitive power. For this reason, democracies must act before cognitive asymmetries are translated into stable forms of domination, and before circumscribed practices become ordinary elements of the political and informational architecture.

Nowadays, contemporary democratic systems are not immune from troubling dynamics. Facial recognition at demonstrations, social media screening in administrative or migration procedures, predictive policing, and intelligence surveillance show that the problem does not concern authoritarian coercion alone. The vulnerability of democracies may emerge in the possibility of internal illiberal drifts: formally elected governments may progressively constrain media independence or delegitimise oversight institutions, exploiting legal instruments and informational technologies without openly resorting to violent repression [37]. Moreover, political actors and interest groups may employ strategies of algorithmic persuasion based on data profiling and psychographic microtargeting [38, 39]. The architecture of digital platforms, structured to maximise attention and engagement, encourages audience segmentation and message personalisation, thereby enabling political content to be adapted to individuals' specific cognitive vulnerabilities. Techniques of digital nudging, targeted disinformation, and the engineering of consent may contribute to shaping citizens' mental representations, reducing their reflective autonomy. If the consent of the people governed is shaped through opaque informational architectures and techniques of cognitive influence, it becomes problematic to determine the extent to which it remains genuinely free.

In this respect, it is necessary to clarify that contemporary cognitive power should not be understood exclusively as state power, nor as a phenomenon confined to authoritarian regimes. The disclosures associated with WikiLeaks and, later, with Edward Snowden revealed the extent to which surveillance practices may also involve liberal democracies, intelligence agencies, and their relations with private technological infrastructures [40]. States, through governments, administrative apparatuses, and intelligence agencies, may rely on the cooperation of private companies which, over recent decades, have acquired an unprecedented capacity to collect data, profile behaviour, organise informational flows, and steer public attention. Major digital platforms are therefore not merely external instruments that may

occasionally be used by political power. They are endogenous infrastructures of the contemporary cognitive environment, since they help define the conditions under which citizens access information, form opinions, and participate in the public sphere. Consequently, distortions of mental freedom do not arise only from authoritarian deviations or from excessive executive power, but may also emerge from the ordinary structures of digital capitalism, grounded in data extraction, algorithmic personalisation, and the monetisation of attention [41]. The alliance between public powers and private platforms can further intensify these dynamics, producing a hybrid concentration of informational power capable of weakening citizens' cognitive autonomy even within liberal democracies.

Looking ahead, these dynamics may be further amplified by the integration of big data, artificial intelligence, and applied neuroscience. The interaction between algorithmic surveillance, informational manipulation, and digital persuasive technologies can generate emerging forms of systemic cognitive pressure capable of affecting attentional processes, memory, emotional regulation, and decisional autonomy. Devices capable of monitoring brain activity, analysing emotional states, and inferring behavioural intentions open unprecedented scenarios of neural surveillance and potential psychological manipulation [42]. Beyond the dimension of physical and informational control, an increasingly sophisticated level of cognitive influence and mental manipulation thus emerges. From a neuroethical perspective, this configuration raises particularly significant questions concerning the practicability of freedom of thought and citizens' decisional autonomy.

Neurorights and the Political Dimension of Mental Freedom

In recent years, a weakening of democratic systems has occurred on a global scale. A large proportion of the world's population now lives under authoritarian regimes, and even long-established democracies are exhibiting signs of crisis [43]. Moreover, mechanisms of digital control and mental manipulation are no longer confined to traditionally authoritarian contexts; rather, they are increasingly permeating contemporary political systems in a cross-cutting manner [41]. It is against this backdrop that the need has emerged to develop innovative conceptual tools capable of safeguarding the most intimate dimension of individual freedom and of protecting the mental sphere of citizens from interference by intrusive state powers marked by growing authoritarian tendencies. The aim is to ensure effective protection of inner autonomy, beyond the traditional forms of rights protection.

Within this framework, the debate on neurorights has emerged as one of the principal contemporary efforts to address the ethical, legal, and social implications of technologies capable of interacting – directly or indirectly – with mental processes. The term "neurorights" was introduced by Ienca and Andorno with the aim of systematically confronting the challenges posed by the development of neurotechnologies in terms of high-level normative principles, such as rights, prerogatives, and duties associated with the cognitive sphere [44]. From this perspective, neurorights may be defined as principles of freedom relating to the cerebral and mental domain of the individual, that is, as fundamental normative rules intended to ensure that technological innovation does not compromise individual autonomy, deliberative capacity, and citizens' self-determination.

The fundamental theoretical reference point is the concept of cognitive liberty, which is rooted in the tradition of political liberalism and in the idea that the human mind constitutes the core of personal freedom and self-determination. As early as the nineteenth century, John Stuart Mill maintained that each individual is sovereign over his or her own body and mind, thereby laying the foundations for a conception of citizenship grounded in mental autonomy [45]. In the contemporary era, Sententia defined such liberty as the right to control one's own consciousness and thought processes, while Bublitz has interpreted it as a form of cognitive self-determination [46, 47]. From a closely related perspective, Yuste and colleagues have argued for the existence of a "right to agency", understood as freedom of thought and the capacity to choose one's actions autonomously [48].

Cognitive liberty constitutes the necessary substratum for every other freedom, insofar as it underpins moral and political subjectivity. In line with this approach, Boire and Sententia have argued that cognitive liberty should not be interpreted merely as a neurophilosophical category or as a moral ideal, but rather as a genuine fundamental right, deserving explicit legal recognition [47, 49]. Ienca and Andorno

have likewise maintained that cognitive liberty protects the sphere of thought prior to any externalisation through speech, writing, or conduct [44]. As such, it precedes any other liberty and stands in a complementary relationship to notions such as freedom of expression, freedom of the press, and freedom of assembly.

A second fundamental dimension, closely linked to mental freedom and the capacity for autonomous action, is the protection of privacy. The profound relationship between privacy and freedom of thought was already recognised by Bury, who observed that a person can never be prevented from thinking what he wishes, so long as he conceals what he thinks [50]. Traditionally, privacy has been conceptualised as "the right to be let alone", a notion subsequently expanded by Westin into that of "information privacy", understood as control over data relating to oneself [51, 52]. In the present context, the challenges extend beyond the analysis of primary neural data, such as recordings of brain activity, and concern above all the inferences drawn from secondary behavioural data. In this respect, Yuste and colleagues have highlighted how an extraordinary volume of personal information can already be obtained through social media profiling practices, algorithmic microtargeting, and predictive analysis supported by artificial intelligence [53].

In the light of these developments, Ienca and Andorno have proposed an evolutionary reinterpretation of the right to privacy, arguing for the need to recognise a genuine "right to mental privacy", aimed at protecting individuals from unauthorised access to their cognitive, emotional, and intentional states [44]. Moreover, they situate their proposal within the broader framework of international human rights law, referring to the United Nations Universal Declaration of Human Rights (1948), the Charter of Fundamental Rights of the European Union (2000), and UNESCO's Universal Declaration on Bioethics and Human Rights (2005). While acknowledging the fundamental value of these instruments, the authors contend that human rights must be "updated" in order to respond adequately to the challenges posed by emerging technologies. According to the two scholars, the possibilities opened up by technological developments and their application across multiple domains of human life, including the political sphere, render a reconceptualisation of certain fundamental rights – and, where necessary, the creation of new rights – inevitable, so as to protect individuals from potential cognitive, psychological, and identity-related harms.

It is important to emphasise that the debate on neurorights does not concern exclusively experimental technologies or applications confined to clinical or laboratory settings; rather, it extends to digital tools that are now fully integrated into citizens' everyday lives, including those employed for political communication, civic participation, and the organisation of collective mobilisation, such as social media platforms. In this respect, dynamics of mental influence and cognitive shaping already operate directly within the public sphere and the institutional domain, and may, as noted above, also be deployed for repressive or authoritarian purposes. Consistently with this broader perspective, Ienca and Vayena have proposed an expansive definition of neurorights, encompassing alterations of mental states induced by technologies that are not strictly neural, such as social media and online manipulation practices, irrespective of whether such interventions result in enhancement, diminution, or no change in brain function [54].

More recently, the debate on neurorights has developed significantly within the scientific community, particularly in the journal *Neuroethics*. One of the central issues to emerge in this discussion concerns the problem of normative demarcation: under what circumstances can a technological alteration of mental processes be regarded as a legitimate component of communication and social interaction, and in which cases, by contrast, should it be deemed impermissible? Some scholars have invoked the concept of manipulation, arguing that a mental intervention violates cognitive liberty insofar as it assumes a manipulative character [55]. McCarthy-Jones maintains that the right to freedom of thought encompasses the right not to be subjected to manipulation in one's thinking [56]. From a more restrictive perspective, Keese and Leiser contend that only a limited set of manipulative practices can plausibly amount to such a violation – namely, those in which mental autonomy is threatened or compromised in a sufficiently "global" manner [57].

Istace, by contrast, has proposed moving beyond an approach centred on the notion of manipulation, developing criteria grounded in the protection of mental autonomy [58]. According to Istace, an intervention should be regarded as morally wrongful when

an agent intentionally undermines a subject's mental autonomy by bypassing, distorting, or disabling the mental capacities that make it possible. Improper interventions profoundly affect an individual's fundamental interests; among the most salient examples are interferences with political opinions. Istace further emphasises that the analysis of the mind's vulnerability should not be confined to emerging neurotechnologies. Numerous social and communicative practices are capable of interfering with individuals' privacy, freedom, and mental integrity, including persuasive conversations, ideologically oriented educational programmes, digital marketing, disinformation campaigns, and forms of governmental indoctrination. Such practices represent varying degrees of threat to freedom of thought, particularly to the right not to be subjected to impermissible alterations of one's cognitive processes.

The various forms of technologically mediated intervention may be situated along a continuum of severity, depending on their impact upon the subject's autonomy. An intervention qualifies as wrongful when an agent intentionally undermines mental autonomy by bypassing, distorting, or neutralising the cognitive capacities that enable an individual to form judgements, assess information, and make decisions independently. These theoretical approaches prove valuable in understanding the strategies of cognitive influence adopted by governments exhibiting authoritarian tendencies. For instance, in the context of the repressed mobilisations in Iran, the systematic use of digital propaganda, targeted disinformation, and algorithmic surveillance has contributed to the emergence of dynamics of self-censorship and cognitive disempowerment, progressively eroding citizens' capacity to oppose the actions of the regime.

A further significant contribution to the debate on neurorights is offered by Ligthart, who explores the normative relevance of the concept of exploitation, distinct from yet complementary to those of manipulation and wrongful interference [59]. In this context, attention shifts not only to the modalities through which a subject is influenced, or to the effects upon the individual, but also to the power structures within which such influence is embedded.

Feinberg defines exploitation as a form of instrumental use of a person for one's own ends that is, in some respect, wrongful [60]. Exploitation is typically detrimental, as it entails an asymmetrical relationship in which one party derives advantage at the expense of another. From a similar perspective, Wood distinguishes exploitation from mere use, arguing that it occurs when a person is incorporated as a means within another's plan, in circumstances where the exploiter exercises a form of domination or control [61]. What characterises exploitation, therefore, is not use as such, but the presence of a structural relationship of subordination, in which one party loses control over the conditions of his or her own action.

Ligthart draws upon these elements in order to apply them to the context of technologies and practices of mental influence [59]. He demonstrates how wrongful influence may be interpreted as a specific form of exploitation, insofar as it entails domination over and control of another's cognitive capacities. Such domination may be exercised not only by private actors – such as companies engaged in neurotechnological experimentation, digital platforms, or predictive marketing practices – but also by public actors, particularly governments seeking to subordinate and discipline the population. As Wood observes, exploitation is degrading for those who are subjected to it, since it deprives them of freedom, a fundamental component of human dignity. In this sense, the erosion of mental autonomy does not merely constitute a functional loss, but amounts to a genuine injury to moral personality. Dynamics of this kind may be observed in contexts of repressed mobilisation, such as the failed uprisings in Iran, Belarus, or Myanmar, where propaganda, digital censorship, and psychological intimidation have converged in practice.

The concept of exploitation has the advantage of possessing well-established legal relevance within international law [62–64]. In the international legal tradition, the exploitation of persons and their bodies has been identified as the principal moral wrong underlying the absolute prohibition of slavery and extreme forms of subordination. In light of this, Ligthart suggests that a comparable normative framework could be extended to the exploitation of minds [59]. If bodily and labour exploitation are deemed impermissible on the grounds that they destroy personal autonomy and dignity, then cognitive exploitation may likewise constitute a moral wrong sufficiently serious to justify the absolute prohibition of certain forms of mental interference. From this perspective, some forms of technological intervention upon the mind would not merely be subject to

regulation or limitation, but ought to be regarded as intrinsically illegitimate. In particular, this category would encompass practices that systematically aim to deplete deliberative capacities, to entrench conditions of cognitive dependence, or to render individuals structurally vulnerable to political power.

Recourse to the concept of exploitation thus serves to strengthen the theoretical foundations of neurorights by integrating the individual dimension of autonomy with a broader analysis of power relations. Whereas approaches centred on manipulation or wrongful interference focus primarily on discrete acts, the notion of exploitation captures the systemic, enduring, and structural character of many practices of cognitive influence. It follows that, in order to be fully effective, neurorights should not be confined to protecting individuals from isolated instances of mental violation, but should also seek to counter the social and political conditions that make the exploitation of cognitive faculties possible. Only in this way can autonomy, dignity, and democratic citizenship be genuinely safeguarded in the age of technologies of influence.

Integrating Neurorights into Habeas Mentem. Toward a Systemic Framework of Protection

In the past, one of the fundamental political achievements was the principle of Habeas Corpus, which literally means "that you have the body". It constituted a right ensuring that individuals could not be imprisoned without just cause, thereby providing protection against abuses of physical power by the State. However, Habeas Corpus was not merely a collection of isolated rights; rather, it was the expression of a principle of systemic significance, aimed at redefining the relationship between the individual and authority. From its formalisation in the Magna Carta of 1215 and, subsequently, in the Habeas Corpus Act of 1679, this principle contributed to limiting arbitrary power by imposing the obligation to provide public justification for any form of detention [16, 65]. In so doing, it introduced a mechanism of balance between executive and judicial power, strengthening the role of guarantor institutions and laying the foundations of constitutionalism and the modern rule of law [66]. Habeas Corpus thus affirmed, for the first time in a systematic manner, that the citizen's body could not be subjected to State control without transparent procedures, subject to verification and independent oversight. Although the individual claim to liberty was central, Habeas Corpus also functioned as a structural principle for the limitation of power, initiating the subjection of authority to constraints and checks and balances [67].

Today, however, with the advent of advanced technologies such as artificial intelligence and neurotechnologies, the power exercised by regimes is no longer confined to physical control, but extends to psychological control. In the contemporary context, power is no longer exercised solely through physical coercion or territorial domination, but through the capacity to collect, analyse, and steer individuals' mental, emotional, and decision-making processes. Governments, together with technology companies, now possess tools capable of systematically influencing citizens' attention, preferences, beliefs, and political behaviour. This development risks generating new forms of cognitive inequality, whereby certain actors concentrate disproportionate power over the formation of collective opinions and wills. In the face of the risk of despotic drift, and in light of the view that neurorights are necessary yet not sufficient, it is urgent to explore which principles, rules, and institutions – conceived in the most systemic and comprehensive sense – may assist citizens in defending their freedom and self-determination when these are threatened within the digital sphere.

For this reason, there is a perceived need for a new general principle, which may be termed Habeas Mentem (literally, "that you have the mind"), aimed at protecting individuals' mental freedom and cognitive integrity. This ethical–political principle must guarantee citizens a fundamental right to mental liberty, prohibiting unauthorised interventions or manipulations of their cognitive activity. However, Habeas Mentem cannot be understood exclusively as a set of neurorights, even if these constitute its core. Rights such as cognitive liberty, mental privacy, the right to psychological integrity, the continuity of personal identity, and, more broadly, decisional self-determination represent indispensable instruments for the protection of subjectivity in the digital age. It is essential to prevent the unlawful collection of neural data, the manipulative use of predictive algorithms, and the imposition of invasive technologies of surveillance and monitoring. To reduce Habeas Mentem, however, to a

catalogue of individual rights would risk underestimating its political and institutional significance. The formal recognition of new rights, such as neurorights, may prove insufficient to counter effectively the structural power dynamics accompanying the development of neurotechnologies. Rather, in a manner analogous to Habeas Corpus, Habeas Mentem must be interpreted expansively as a principle for the balancing of power, designed to redress the cognitive and material asymmetries that arise between citizens, States, and major technological actors.

In this sense, Habeas Mentem should assume the function of a regulative principle of cognitive equilibrium, aimed at preventing concentrations of power from translating into forms of mental domination. It does not merely protect individuals against isolated violations, but seeks to structure institutionally the relationships between persons and cognitive powers, by imposing limits, oversight, and obligations of transparency. For example, it could entail the establishment of independent authorities, mechanisms of public supervision, and instruments of legal accountability capable of monitoring the exercise of mental power. It is therefore urgent to identify which competences, rules, and institutions may assist citizens in defending their freedom in the age of artificial intelligence, when it is placed under threat. The principle of Habeas Menten may be articulated across multiple levels – legal, political, social and educational – through specific rights, systemic corrections, and dedicated institutions (Fig. 1).

This articulation should not be understood as a mere list of heterogeneous measures. Rather, it reflects the dual nature of Habeas Mentem. On the one hand, the principle performs a protective function (which, as will be discussed more extensively in the concluding discussion, is close to Berlin's conception of "negative liberty"), since it aims to defend individuals against illegitimate forms of cognitive interference, informational exploitation, neural and behavioural surveillance, and algorithmic manipulation. On the other hand, it performs an enabling function

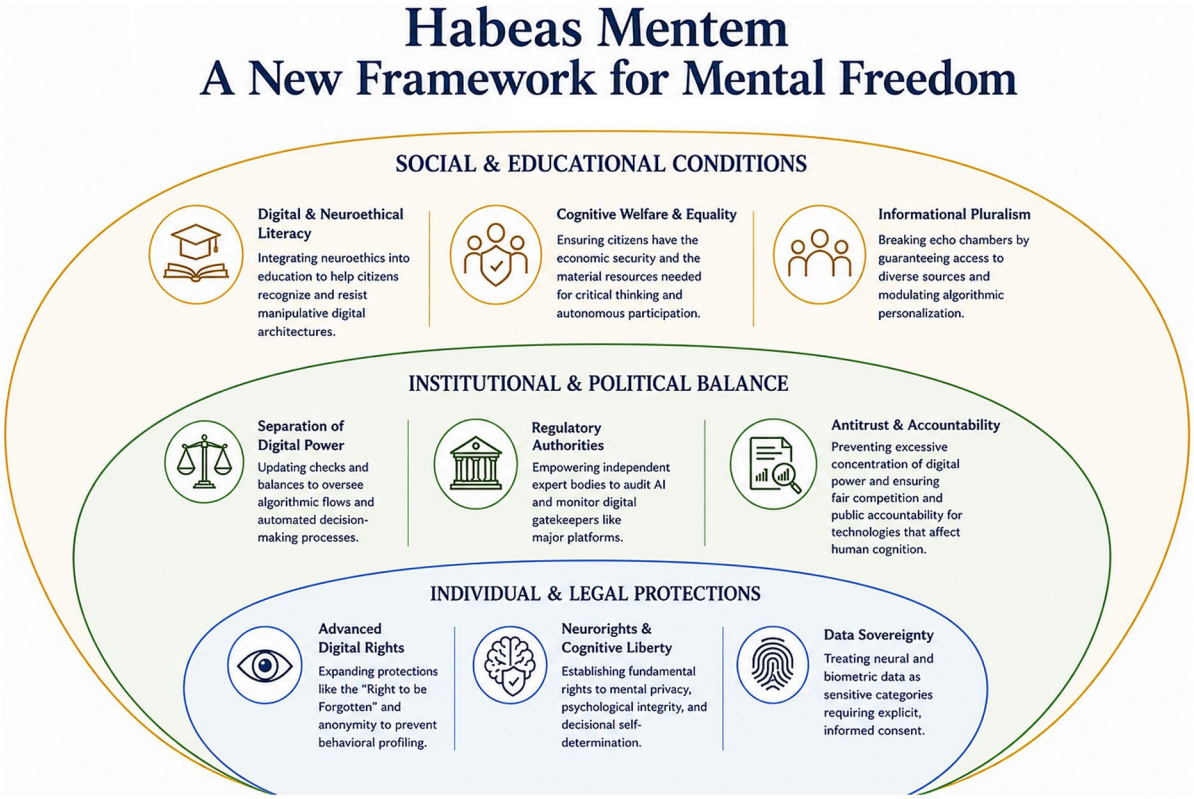

Fig. 1 The articulation of Habeas Mentem in the digital age

(close to the conception of "positive liberty"), since mental freedom also depends on the social, institutional, and epistemic conditions that allow citizens to form autonomous judgements, access pluralistic information, understand digital environments, and participate meaningfully in democratic life. The subsections that follow should therefore be read as interdependent levels. Individual and digital rights define the defensive core of the framework. The separation of digital power, platform regulation, and antitrust constitute its institutional dimension. Education, cognitive welfare, and informational pluralism express its social and enabling dimension.

Neurorights and Associated Digital Rights

Habeas Mentem signifies, first and foremost, a comprehensive updating of the human rights system, aimed at safeguarding individual freedom in the management of informational elements conveyed through digital technologies. This entails recognising the right to access information autonomously, to protect one's sensitive data, to communicate one's views freely, and to resist conditioning exerted by political and economic power through technological instruments.

Within this framework, neurorights do not occupy a separate sphere from traditional digital rights; rather, they represent their most advanced frontier. They seek to protect the deepest level of informational experience, namely that at which data, communications, and digital interactions translate into cognitive, emotional, and decision-making processes. Digital rights may therefore be interpreted as a concrete and operational articulation of neurorights within the context of the Internet, which remains the principal arena in which dynamics of cognitive influence, persuasion, and manipulation are exercised [68]. Flores Filho and Firmo have already suggested that the full implementation of neurorights should find a corresponding expression in the regulation of the Internet [69]. The protection of mental freedom presupposes, indeed, the proper governance of the informational environments within which the mind operates. As these environments are now largely structured by digital platforms, recommendation algorithms, and profiling systems, the protection of personal data, digital traces, and communicative flows becomes an essential condition. As Brown argues, genuine safeguards for the mental sphere require an expansion of the protective scope beyond devices directly targeting the brain, so as to encompass the full spectrum of risks [70]. In the digital age, without adequate regulation of the online informational space, neurorights understood in a strict sense risk remaining ineffective.

The hypothesis of a systematic articulation of rights connected to information and communication technologies continues to encounter a degree of resistance, as some fear that it may conflict with the libertarian tradition of the Internet. To this position, however, it may be objected – following Rodotà – that institutional innovation in the digital sphere does not seek to curtail the freedoms fostered by technological expansion, but rather to render them effective and sustainable over time, preventing their erosion through asymmetries of power between users and governments or major technological actors [71].

A first digital right is that of access to the Internet, understood not merely as the formal possibility of connection, but also as the overcoming of the digital divide in terms of infrastructure, skills, and availability of devices. Effective access to online informational environments constitutes a preliminary condition for the exercise of cognitive autonomy and democratic participation. A second fundamental right is that to privacy and the protection of personal data, an issue directly addressed in discussions on neurorights. As Yuste warns, gaps in existing regulatory frameworks risk permitting the almost unrestricted decoding and commercialisation of neural data [72]. Nevertheless, limitations on the collection and use of behavioural data are already present within Internet law: the online dissemination of certain categories of sensitive data, for instance, is prohibited within the European Union. Moreover, this right must guarantee individuals control over information relating to them and over the manner in which such information is collected, processed, and used. It also entails the possibility of giving informed consent, of knowing the purposes of data processing, and of opposing invasive or discriminatory practices. A noteworthy example is the European General Data Protection Regulation (GDPR), which introduces key principles such as data minimisation and purpose limitation, thereby offering a first line of defence against the exploitation of cognitive processes. Specifically, the Regulation classifies biometric data and data concerning health as "special categories of data", requiring enhanced safeguards and explicit consent. The protection of personal data thus constitutes an essential

condition for preserving the private sphere, preventing forms of informational manipulation, and limiting the asymmetrical power exercised by digital platforms and institutional actors. In this respect, privacy represents a central safeguard of personal autonomy and of the freedom to form one's thoughts.

A third element is the right to be forgotten, namely the individual's entitlement to request the removal, de-indexing, or contextualisation of personal information available online. As Jones emphasises, this right performs a valuable anti-authoritarian function, insofar as it limits the ability of public and private actors to accumulate permanent data conducive to cognitive profiling and behavioural prediction [73]. In this sense, the right to be forgotten contributes directly to the protection of mental identity and psychological self-determination. A further essential right is that to anonymity, which enables users to communicate, access information, and participate in public debate without being systematically identified and tracked. Through encryption tools and privacy-enhancing systems, anonymity reduces the risk of surveillance, repression, and self-censorship, thereby preserving the cognitive space necessary for intellectual experimentation, dissent, and the autonomous formation of opinions.

More generally, as Gary T. Marx observes, it is essential to engage in systematic scrutiny of the purposes, modalities, and individual and collective consequences of data collection by authorities [74]. In liberal-democratic countries, albeit with interpretative variations, it is generally maintained that data should be collected fairly, for legitimate and proportionate purposes, and that they should be accurate, relevant, up to date, and accessible to data subjects for correction or contestation. They should not exceed their declared purposes, be retained longer than necessary, nor be disclosed to third parties outside narrowly justified exceptions. By contrast, in countries characterised by authoritarian or repressive tendencies, data collection ceases to function as an administrative service and instead becomes an instrument of pervasive social control.

Separation of Digital Power and a New System of Checks

The principle of Habeas Corpus constituted the first formulation of a proto-liberal political thought and contributed to inaugurating a conception of power as a reality constrained by rights, procedures, and oversight. In the centuries that followed, this approach was further elaborated by thinkers such as Locke and Montesquieu, not only through the theory of rights but also through the doctrine of the separation of powers [75, 76].

In the contemporary context, power is no longer concentrated solely within traditional political institutions, but is distributed among States, technological platforms, transnational corporations, and algorithmic systems [77]. Within this scenario, traditional forms of separation of powers may prove inadequate to capture the new dynamics of cognitive and informational domination. Consequently, it becomes necessary to rethink the principle of the separation of powers in a digital key, developing new mechanisms of control. This may entail, for example, the creation of independent institutions tasked with overseeing algorithms, informational flows, and data usage, as well as the introduction of public procedures for the evaluation, auditing, and contestation of automated decision-making [78]. At the same time, the role of public authorities must be strengthened in supervising technologies that affect citizens' cognitive and decisional processes.

In the digital sphere, the concentration of surveillance capacities, data collection, and informational manipulation in the hands of the executive constitutes a structural risk to citizens' freedoms. The traditional distinction between executive, legislative, and judicial power – conceived in order to limit political arbitrariness – must therefore be updated. Within this framework, the legislature is called upon to establish a general regulatory framework setting binding principles concerning algorithmic transparency, data protection, platform accountability, and limits on surveillance. The judiciary, for its part, must perform an essential ex post oversight function, ensuring the possibility of challenging abusive practices, discriminatory automated decisions, and unlawful forms of monitoring. In so doing, it operates as a safeguard against the instrumental use of technology for repressive purposes.

Alongside these traditional powers, independent regulatory authorities may play a complementary role, including data protection agencies and communications oversight bodies [79]. In particular, the opacity and systemic risks associated with artificial

intelligence require a form of governance specifically informed by technological expertise. Although not directly legitimised through popular vote, such institutions are composed of sectoral experts and may perform a crucial function by virtue of their technical competence and relative autonomy from political direction [80]. Equipped with investigative and sanctioning powers, these bodies can contribute to limiting data collection, overseeing digital operators, and protecting citizens. Owing to their autonomy, they may also act in an adaptive and anticipatory manner [81].

In the digital age, it is not only public power that poses a potential threat to freedom, but also private power, especially when concentrated in the hands of a few actors such as major technological platforms. Modern constitutional systems were primarily designed to limit the authority of the State, not to regulate entities such as Google, Meta, or Amazon, whose influence rests upon the control of vast quantities of data, opaque algorithms, global infrastructures, and widespread relations of economic dependence. The development of artificial intelligence likewise tends to follow a highly concentrated trajectory capable of sustaining authoritarian political drift [82]. As Wu observes, the colossal scale of digital platforms generates consequences that extend far beyond the strictly economic sphere [83]. Major technology companies exert significant influence over the public space, shaping the visibility of content and intervening in decision-making processes through lobbying and financial support. Digital laissez-faire has for years overlooked the dangers of predatory mergers, unlimited data accumulation, technological lock-in mechanisms, and privately controlled essential infrastructures. The result has been the formation of a global oligopoly with nearly insurmountable barriers to entry. Competition is a fragile good: without public intervention, it erodes and ultimately disappears [84].

The division of power therefore also entails strengthening antitrust competences and criteria, as well as intensifying oversight of major technology companies [85–87]. Various further proposals warrant consideration. For example, the burden of proof in merger proceedings could be reversed, requiring companies to demonstrate that an acquisition will not harm competition; mergers could be reviewed even years after their completion; and the value of data could be taken into account in assessing acquisitions. It may also be appropriate to revive the tradition of the major twentieth-century antitrust cases that reshaped entire sectors. In the presence of particularly harmful concentrations, authorities might consider measures of functional or corporate separation, for instance by mandating the structural separation of core infrastructures, intermediation services, and commercial activities. Such interventions are designed to prevent a single entity from simultaneously controlling multiple levels of the digital value chain, thereby reducing the potential for abuse. Additional measures could include the introduction of mandatory independent audits of algorithms, the periodic publication of reports on data governance, and the establishment of public registers of major acquisitions and technological partnerships.

In this direction moves the Digital Markets Act (DMA) of 2022, which inaugurates a new phase in the European regulation of digital platforms [88]. The Regulation addresses so-called “gatekeepers”, imposing obligations concerning data portability, prohibitions on self-preferencing, advertising transparency, and the freedom to operate outside dominant platforms. It represents an intervention aimed at fostering a more equitable ecosystem for start-ups and small and medium-sized enterprises, although its effectiveness will depend upon the capacity of European authorities to enforce its provisions rigorously.

Ultimately, the protection of freedom within the digital ecosystem requires a substantial updating of the constitutional logic of checks and balances. Beyond more effectively constraining political power, it is necessary to articulate a multi-level system capable of containing the informational and economic power of major platforms. Dividing power entails distributing competences among legislators, judges, independent technical authorities, and competition bodies, thereby avoiding concentrations of decision-making authority capable of unilaterally shaping citizens’ cognitive processes. Only through an appropriate division of competences, independent oversight, and the effective enforcement of sanctions can informational pluralism, decisional autonomy, and individual liberty be preserved in the digital age.

Education, Cognitive Welfare and High-quality Information

The system of Habeas Mentem cannot rest exclusively upon legal prohibitions or institutional mechanisms

of balance and oversight; it also presupposes the effective capacities of individuals and their genuine opportunity to develop them fully. Mental freedom, indeed, does not consist merely in the absence of external interference, but requires cognitive and material conditions that enable individuals to understand the informational environment, to navigate within it, and to take autonomous decisions. In the absence of such conditions, individuals lacking adequate interpretative tools cannot effectively exercise the rights that the law formally recognises, nor can they sustain renewed forms of democratic balancing.

From a pedagogical perspective, digital education must assume a central role within school curricula, complemented by elements of neuroethics [89, 90]. Educating citizens capable of understanding technologies enhances their ability to recognise risks and to defend themselves against manipulative practices [91]. Yet the Habeas Mentem approach demands more. Without the availability of time, a minimum degree of economic security, and stable access to informational resources, the aspiration to cognitive autonomy remains substantially diminished. Digital competence is not merely a technical skill, but a capacity that presupposes attention, concentration, continuity in learning, and the opportunity for ongoing intellectual development. For this reason, the educational dimension inevitably intersects with questions of redistribution [92, 93]. Where a significant portion of the population is absorbed by employment precarity, informational overload, or dependence upon platforms designed to capture and monetise attention, the full development of the critical competences required for digital citizenship becomes unattainable.

The digitalisation of the economy renders this problem even more evident. Large platforms extract value both from the data generated by users and from the often invisible labour of digital micro-workers, yet without remunerating it adequately [94]. As a result, the economic value produced socially does not return in the form of public cognitive goods. A fairer system of taxation for digital enterprises could therefore support new forms of welfare oriented towards strengthening citizens' cognitive capacities [95]. This would entail not only universal access to digital infrastructures, but also guaranteed time for education and training, public spaces for learning, free continuing education, and adult reskilling programmes.

In this way, the digital divide would be addressed not merely as a lack of connectivity, but as a gap in cognitive competences [96]. On the one hand, operational technical skills are required: understanding how algorithms function, knowing how to configure privacy settings, using software autonomously and, at least at a basic level, being able to modify and adapt technologies. On the other hand, critical and humanistic competences are equally necessary: logical reasoning, recognition of fallacies, source analysis, and an understanding of the rhetorical and cognitive mechanisms underpinning disinformation. Only the integration of technical capacity and critical education enables individuals to identify informational manipulation, recognise fake news, and discern the interests embedded within digital architectures.

The availability of a high-quality informational environment, within which a civic culture may gradually be formed, likewise constitutes a structural condition of Habeas Mentem and a form of democratic lifelong learning. Pluralist theories of democracy presuppose a specific form of citizen autonomy: not merely the freedom to express opinions, but the capacity to form them through effective engagement with alternative perspectives. The stability of democratic order depends upon the existence of a shared cognitive space within which differences of values and interests may be understood, debated, and contested. Deliberation presupposes, in other words, at least a partially common informational foundation [97].

The informational ecosystem of digital platforms, in the age of big data and recommendation algorithms, introduces instead a structural transformation of the public sphere. Algorithmic personalisation of content tends to fragment the experience of reality and to situate individuals within epistemically homogeneous environments. Platforms select and rank content on the basis of users' past behaviour, with the economic objective of maximising time spent and interaction. Within this context, phenomena commonly described as echo chambers emerge: informational environments in which users predominantly encounter content consistent with their prior preferences [98]. This does not amount merely to increased exposure to similar opinions, but to a genuine preselection of cognitive experience itself [99].

From a cognitive perspective, the automated selection of content attenuates systematic exposure to

contradiction and renders cognitive dissonance intermittent – yet, under ordinary conditions, such dissonance performs an essential epistemic function in prompting the revision of beliefs. Secondly, informational redundancy increases, as users receive multiple variations of the same interpretative content: news, commentary and explanations converge upon a single narrative framework, and repetition ultimately reinforces subjective plausibility independently of the epistemic quality of the information. Finally, a stabilisation of beliefs occurs: the internal coherence of the informational environment heightens perceived certainty in one's own views, and knowledge no longer derives from a process of exploration [100]. The consequence is a polarising effect even in the absence of ideological intentionality. Individuals who are initially moderate may gradually adopt more extreme positions, not because they are exposed to more radical arguments, but because informational homogeneity increases subjective confidence in the beliefs available to them [101]. A fragmented and polarised environment is precisely the one most conducive to disinformation [102, 103]. Moreover, these dynamics may in turn be amplified by the development of artificial intelligence, which enables the production of increasingly realistic deepfakes [104].

For Habeas Mentem, the issue concerns the overall epistemic architecture within which individuals acquire information. Effective mental freedom requires the structural possibility of encountering cognitive alterity. Without informational pluralism, there can be no genuine deliberative autonomy, but merely the perception of choosing within a space that has already been preselected. Accordingly, protection cannot be confined to the removal of unlawful content or to the factual verification of news. Systemic conditions are required to guarantee access to diverse and independent sources of information, transparency in the criteria governing algorithmic recommendation, the possibility of deactivating or modulating personalisation, and the promotion of digital public spaces not governed solely by the logic of engagement.

In the absence of such conditions, individuals remain formally free yet cognitively steered: choice remains legally voluntary, but epistemically heterodirected. Nor can democratic participation be fully realised, since it requires not only formal access to the public sphere, but also the cognitive, informational, and material capacities needed to understand issues, evaluate alternatives, and intervene meaningfully in collective decision-making. Consistent with its enabling structure, Habeas Mentem therefore entails not only negative rights against mental intrusion, but also positive guarantees concerning the quality of the environment within which decisions are formed.

Concluding Discussion

Habeas Mentem may be conceived as a principle of integration among rights, rules, and institutions, aimed at ensuring that the cognitive power exercised through technologies respects citizens' freedom. It may be interpreted as a paradigm in which the minds of citizens, much like the body in the Habeas Corpus, cannot be subjected to arbitrariness, whether by the State or by private actors endowed with systemic capacities of influence. Its added value does not lie in replacing existing liberal protections, which remain indispensable, but in extending and reorganising them around the specific problem of cognitive power in the digital age. Neurorights constitute its immediate normative core; yet its deeper function lies in redefining power relations within the digital ecosystem, preserving the cognitive conditions necessary for liberty and collective self-government.

To avoid ambiguity, Habeas Mentem should not be understood as a new right, nor as a sectoral programme of technological regulation. It should rather be conceived as a constitutional-level meta-principle, a normative framework, and a general architecture capable of integrating existing and emerging rights, orienting the distribution of power, and guiding the reform of democratic institutions in the digital age. In this sense, Habeas Mentem does not replace neurorights, but incorporates them as its immediate core, while situating them within a broader architecture of guarantees, rules, and institutions. Indeed, the defence of freedom of thought – and, consequently, of political action – in the digital age cannot be entrusted solely to the formal recognition of new individual rights. Rather, it requires an institutional framework capable of structurally redressing the asymmetries of power generated by the attention economy and the concentration of informational infrastructures.

Habeas Mentem takes shape as a multi-level system of power-balancing, integrating the separation of state powers, the regulation of private enterprises,

antitrust policies, and the autonomous participation of civil society. From this perspective, the protection of cognitive and communicative freedom does not depend upon isolated legal or regulatory instruments, but upon the construction of a pluralistic institutional ecosystem capable of distributing technological power, rendering it transparent, and subjecting it to reciprocal checks.

More specifically, the framework of Habeas Mentem developed in Sect. 3 operates along two complementary dimensions, which may be understood in light of Isaiah Berlin's distinction between negative and positive liberty [105]. The first dimension is that of negative freedom, understood as protection against illegitimate interference in the mental and cognitive sphere. It includes the prohibition of unauthorised access to neural and behavioural data, protection against manipulative algorithmic practices, limits on surveillance, independent oversight mechanisms, and the balancing of powers. From this perspective, Habeas Mentem takes up the defensive function of neurorights and digital rights, protecting individuals from arbitrary cognitive intrusion, informational exploitation, and psychological manipulation.

The second dimension is that of positive freedom, understood as the construction of the material, social, and institutional conditions that make mental autonomy effectively possible. Cognitive freedom cannot be reduced to the mere absence of interference. It also requires access to pluralistic information, critical education, digital literacy, welfare, and material conditions enabling individuals to understand, contest, and participate within complex technological environments. From this perspective, educational policies, the taxation of digital platforms, the production of public cognitive goods, informational pluralism, and democratic participation are not external or supplementary elements of Habeas Mentem, but constitutive parts of its systemic architecture, aimed at correcting the material dimensions of inequality in the digital world that risk weakening citizens' positive freedom.

These components should not be understood as a mere juxtaposition of heterogeneous measures, but as different levels of the same general normative architecture. Indeed, no single measure, in this theoretical framework, is sufficient. The legal protection of mental data, the regulation of algorithms, antitrust policies, platform transparency, critical education, and the development of public cognitive goods operate at distinct yet interdependent levels. Only their parallel and coordinated implementation can prevent mental freedom from remaining a merely formal guarantee. Habeas Mentem thus assumes a systemic and enabling structure, analogous to that of modern constitutional arrangements in which fundamental rights and the organisation of powers mutually reinforce one another. In this way, mental freedom is conceived not merely as immunity from interference, but as the institutional outcome of a dynamic equilibrium among power, knowledge, and democratic participation.

The specific contribution of the article therefore lies in shifting the centre of gravity of the debate from neurorights understood as defensive individual rights to Habeas Mentem understood as a normative principle, a constitutional meta-principle and a systemic framework for rebalancing cognitive power. The originality of the proposal does not lie merely in the addition of new neuro-rights concerning the digital sphere, but in the connection between individual protections, institutional limits, platform regulation, informational pluralism, critical education, and the social conditions of mental freedom. What Habeas Mentem adds, then, is not another item to the existing liberal vocabulary, but a principle of coordination capable of explaining why these diverse protections must be treated as mutually dependent elements of a single social, political and legal project.

In conclusion, the emergence of pervasive communication technologies and the imminent development of neurotechnologies require a structural reconsideration of the guarantees of liberty within the digital sphere. The expansion of technical capacities to acquire, infer, and modulate cognitive processes transforms the human mind into a newly relevant juridical space, exposed to forms of power not reducible to traditional models of physical or economic coercion. In this context, Habeas Mentem emerges as a systemic regulatory principle aimed at preventing technological power from evolving into cognitive domination. Its purpose is therefore not only remedial, but also constitutional and anticipatory, since it seeks to shape the institutional environment before cognitive asymmetries become normalised structures of dependence. In light of the impending integration of digital technologies and neurocognitive systems, the anticipatory adoption of ethical and legal safeguards becomes a constitutional necessity rather than

a matter of prudence alone. Without such pre-emptive intervention, neurotechnologies risk consolidating new forms of invisible subordination; with it, they may instead be oriented towards emancipatory ends.

Acknowledgements This article was developed during a period of visiting research on the project "Addressing Fake News: Democratic Procedures for Digital Communication" at Pompeu Fabra University, Barcelona, Spain, supported by the Department of Humanities and Cultural Heritage of the University of Udine, Italy. An earlier version of the paper was presented under the title "From Digital Repression by Authoritarian Regimes to the Need for a Habeas Mentem" at the 15th Braga Meetings on Ethics and Political Philosophy, University of Minho, Portugal. I would like to thank the colleagues who contributed with their comments to strengthening the theoretical proposal presented in this article.

Author Contribution All parts of the article were conceived and written by Gabriele Giacomini.

Funding Open access funding provided by Università degli Studi di Udine within the CRUI-CARE Agreement.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing interest The authors declare no competing interests.

References

1. Kendall-Taylor, A., E. Frantz, and J. Wright. 2020. The digital dictators: How technology strengthens autocracy. *Foreign Affairs* 99:103.
2. Keremoğlu, E., and N. B. Weidmann. 2020. How dictators control the internet: A review essay. *Comparative Political Studies* 53 (10–11): 1690–1703.
3. Schlumberger, O., M. Edel, A. Maati, and K. Saglam. 2024. How authoritarianism transforms: A framework for the study of digital dictatorship. *Government and Opposition* 59 (3): 761–783.
4. Deibert, R. J. 2022. Subversion inc: The age of private espionage. *Journal of Democracy* 33 (2): 28–44.
5. Feldstein, S. 2019. The road to digital unfreedom: How artificial intelligence is reshaping repression. *Journal of Democracy* 30 (1): 40–52.
6. Giordano, J., ed. 2014. *Neurotechnology in national security and defense: Practical considerations, neuroethical concerns*. Boca Raton: CRC Press.
7. Saheb, T. 2023. Ethically contentious aspects of artificial intelligence surveillance: A social science perspective. *AI and Ethics* 3 (2): 369–379.
8. Giacomini, G. 2022. *The arduous road to revolution: Resisting authoritarian regimes in the digital communication age*. Milan: Mimesis.
9. Chang, C. C., and T. H. Lin. 2020. Autocracy login: Internet censorship and civil society in the digital age. *Democratization* 27 (5): 874–895.
10. Earl, J., T. V. Maher, and J. Pan. 2022. The digital repression of social movements, protest, and activism: A synthetic review. *Science Advances* 8 (10): eabl8198.
11. Garbe, L., and S. F. Maerz. 2026. The rise of authoritarian informationalism: Escalating surveillance, manipulation, and control. *Democratization* 33 (1): 1–18.
12. Maerz, S. F. 2026. How practices of digital authoritarianism harm democracy. *Democratization* 33 (1): 163–190.
13. Hertz, N. 2023. Neurorights–do we need new human rights? A reconsideration of the right to freedom of thought. *Neuroethics* 16 (1): 5.
14. Ienca, M. 2021. On neurorights. *Frontiers in Human Neuroscience* 15:701258.
15. Muñoz, J. M., and J. Á. Marinaro. 2023. Neurorights as reconceptualized human rights. *Frontiers in Political Science* 5:1322922.
16. Wright, C. A. 1983. Habeas corpus: Its history and its future. *Michigan Law Review* 81 (4): 802–810.
17. Locke, J. 1690. *Two treatises of government*. London: Tegg.
18. Hasebe, Y. 2001. The force to rule, the force to resist: The case of John Locke. *The Sociology of Law* 54:116–129.
19. Seliger, M. 1963. Locke's theory of revolutionary action. *Western Political Quarterly* 16 (3): 548–568.
20. Castells, M. 2012. *Networks of outrage and hope: Social movements in the internet age*. Cambridge: Polity Press.
21. Korany, B., and R. El-Mahdi, eds. 2012. *Arab spring in Egypt: Revolution and beyond*. Cairo: American University in Cairo Press.
22. Aouragh, M., and A. Alexander. 2011. The Arab springl the Egyptian experience: Sense and nonsense of the internet revolution. *International Journal of Communication* 5:15–15.
23. Lynch, M. 2011. After Egypt: The limits and promise of online challenges to the authoritarian Arab state. *Perspectives on Politics* 9 (2): 301–310.

24. Abdelrahman, M. 2011. The transnational and the local: Egyptian activists and transnational protest networks. *British Journal of Middle Eastern Studies* 38 (3): 407–424.
25. Madrigal, A. (2011). "Egyptian Activists' Action Plan." The Atlantic, January 27.
26. Bashiriyeh, H. 2010. Counter-revolution and revolt in Iran: An interview with Iranian political scientist Hossein Bashiriyeh. *Constellations* 17:61–77.
27. Morozov, E. 2009. Iran: Downside to the "Twitter revolution." *Dissent* 56 (4): 10–14.
28. Michaelsen, M. 2018. Exit and voice in a digital age: Iran's exiled activists and the authoritarian state. *Globalizations* 15 (2): 248–264.
29. Conduit, D. 2026. The political economy of digital authoritarianism: Evidence from the Iranian regime's implementation of technology. *Democratization* 33 (1): 62–84.
30. Bashirov, G., S. Akbarzadeh, I. Yilmaz, and Z. S. Ahmed 2025. Diffusion of digital authoritarian practices in China's neighbourhood: the cases of Iran and Pakistan. *Democratization*, 1–24.
31. Akbarzadeh, S., A. Naeni, G. Bashirov, and I. Yilmaz. 2025. The web of big lies: State-sponsored disinformation in Iran. *Contemporary Politics* 31 (2): 328–347.
32. Eichhorn, K., and E. Linhart. 2023. Election-related internet-shutdowns in autocracies and hybrid regimes. *Journal of Elections, Public Opinion & Parties* 33 (4): 705–725.
33. Gohdes, A. R. 2015. Pulling the plug: Network disruptions and violence in civil conflict. *Journal of Peace Research* 52 (3): 352–367.
34. Kostka, G., L. Steinacker, and M. Meckel. 2021. Between security and convenience: Facial recognition technology in the eyes of citizens in China, Germany, the United Kingdom, and the United States. *Public Understanding of Science* 30 (6): 671–690.
35. Howard, P., F. Lin, and V. Tuzov. 2023. Computational propaganda: Concepts, methods, and challenges. *Communication and the Public* 8 (2): 47–53.
36. Guriev, S., and D. Treisman. 2019. Informational autocrats. *Journal of Economic Perspectives* 33 (4): 100–127.
37. Powell, Laura C. 2022. The good: the bad, and the ugly: Black lives matter protests, the January 6th insurrection, and facial recognition technology as admissible evidence. *American University Law Review* 72 (277): 335.
38. Starbird, K., R. DiResta, and M. DeButts. 2023. Influence and improvisation: Participatory disinformation during the 2020 US election. *Social Media+ Society* 9 (2): 20563051231177944.
39. Woolley, S. C. 2022. Digital propaganda: The power of influencers. *Journal of Democracy* 33 (3): 115–129.
40. Lyon, D. 2014. Surveillance, Snowden, and big data. *Big Data & Society* 1 (2): 2053951714541861.
41. Zuboff, S. 2019. *The age of surveillance capitalism*. London: Profile Books.
42. Dunagan, J. F. 2004. Neuro-futures: The brain, politics and power. *Journal of Futures Studies* 9 (2): 1–18.
43. Erdmann, G., and M. Kneuer, eds. 2013. *Regression of democracy?* Berlin: Springer.
44. Ienca, M., and R. Andorno. 2017. Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy* 13 (1): 5.
45. Mill, J. S. 1859. *Essay on liberty*. London: John W Parker and Son.
46. Bublitz, J.-C. 2013. "My mind is mine!? Cognitive liberty as a legal concept,." In *Cognitive enhancement*, ed. H. Franke, 233–264. Berlin: Springer.
47. Sententia, W. 2004. Neuroethical considerations: Cognitive liberty and converging technologies for improving human cognition. *Annals of the New York Academy of Sciences* 1013 (1): 221–228.
48. Yuste, R., J. Genser, and S. Herrmann. 2021. It's time for neuro-rights. *Horizons* 18:154–164.
49. Boire, R. G. 2001. On cognitive liberty. *Journal of Cognitive Liberties* 2:7–22.
50. Bury, J. 1913. *A history of freedom of thought*. Alburgh: Facsimile Publisher.
51. Brandeis, L., and S. Warren. 1890. The right to privacy. *Harvard Law Review* 4:193–220.
52. Westin, A. F. 1968. Privacy and freedom. *Washington and Lee Law Review* 25:166.
53. Yuste, R., S. Goering, G. Bi, J. M. Carmena, A. Carter, J. J. Fins, et al. 2017. Four ethical priorities for neurotechnologies and AI. *Nature News* 551:159–163.
54. Ienca, M., and E. Vayena. 2018. *Cambridge Analytica and Online Manipulation. Scientific American*. Basingstoke: Springer Nature.
55. Teo, S. A. 2024. How to think about freedom of thought (and opinion) in the age of AI. *Computer Law & Security Review* 53:105969.
56. McCarthy-Jones, S. 2019. The autonomous mind: The right to freedom of thought in the twenty-first century. *Frontiers in Artificial Intelligence* 2:1–17.
57. Keese, N., and M. Leiser. 2025. Online manipulation as a potential interference with the right to freedom of thought. In *The Cambridge Handbook for Freedom of Thought*, ed. P. O'Callaghan and B. Shiner. Cambridge: CUP.
58. Istace, T. 2025. Exploring the essence of the freedom of thought. *Neuroethics* 18:10.
59. Ligthart, S. 2025. Enslaving minds: On freedom of thought and the exploitation of mental vulnerabilities. *Neuroethics* 18:48.
60. Feinberg, J. 1990. *The moral limits of the criminal law: Harmless wrongdoing*. Oxford: OUP.
61. Wood, W. 2014. Coercion, manipulation, exploitation. In Manipulation: Theory and Practice, Coons, C., and Weber, M. (eds.). Oxford: OUP.
62. Allain, J. 2013. *Slavery in international law: Of human exploitation and trafficking*. Leiden: Nijhoff.
63. Jovanovic, M. 2020. The essence of slavery: Exploitation in human rights law. *Human Rights Law Review* 20:674–703.
64. Marks, S. 2008. Exploitation as an international legal concept. In *International Law on the Left*, ed. S. Marks. Cambridge: CUP.
65. McFeeley, N. D. 1976. The historical development of haberas corpus. *Southwestern Law Review* 30:585–598.

66. Tabak, R. J. 1995. Habeas corpus as a crucial protector of constitutional rights: A tribute which may also be a eulogy. *Seton Hall Law Review* 26:1477–1492.
67. Wert, J. J. 2023. *Habeas corpus in America: The politics of individual rights*. Lawrence: University Press of Kansas.
68. Mathiesen, Kay. 2014. Human rights for the digital age. *Journal of Mass Media Ethics* 29 (1): 2–18.
69. Flores Filho, E. G. J., and M. C. Firmo. 2023. Human dignity and neurorights in the digital age. *Brazilian Journal of Law, Technology and Innovation* 1 (2): 87–107.
70. Brown, C. M. L. 2024. Neurorights, mental privacy, and mind reading. *Neuroethics* 17 (2): 34.
71. Rodotà, S. 2021. *Tecnologie e diritti*. Bologna: Il Mulino.
72. Yuste, R. 2023. Advocating for neurodata privacy and neurotechnology regulation. *Nature Protocols* 18 (10): 2869–2875.
73. Jones, M. L. 2018. *Ctrl+ z: The right to be forgotten*. New York: New York University Press.
74. Marx, G. T. 1998. Ethics for the new surveillance. *Information Society* 14 (3): 171–185.
75. Duyen, N. T. T. 2022. Separation of powers in John Locke's political philosophy. *Synesis* 14 (1): 1–15.
76. Krause, S. 2000. The spirit of separate powers in Montesquieu. *Review of Politics* 62 (2): 231–265.
77. Durand, C. 2025. *How Silicon Valley unleashed technofeudalism*. Verso Books.
78. Fink, K. 2018. Opening the government's black boxes: Freedom of information and algorithmic accountability. *Information, Communication & Society* 21 (10): 1453–1471.
79. Cohen, J. E. 2016. The regulatory state in the information age. *Theoretical Inquiries in Law* 17 (2): 369–414.
80. Greenleaf, G. 2012. Independence of data privacy authorities: International standards. *Computer Law & Security Review* 28 (1): 3–13.
81. Currie, W. L., J. M. Leimeister, D. Schlagwein, and L. Willcocks. 2025. Rethinking technology regulation in the age of AI risks. *Journal of Information Technology* 40 (3): 236–245.
82. Coeckelbergh, M. 2026. Technofascism: AI, Big Tech, and the new authoritarianism. *AI & Society*, 1–14.
83. Wu, T. 2020. *The curse of bigness*. London: Atlantic Books.
84. Kanter, J. S. 2023. Digital markets and 'trends towards concentration.' *Journal of Antitrust Enforcement* 11 (2): 143–148.
85. Geradin, D., and D. Katsifis. 2022. Strengthening effective antitrust enforcement in digital platform markets. *European Competition Journal* 18 (2): 356–405.
86. Lancieri, F., and C. M. S. Pereira Neto. 2022. Designing remedies for digital markets: The interplay between antitrust and regulation. *Journal of Competition Law & Economics* 18 (3): 613–669.
87. Newman, J. M. 2019. Antitrust in digital markets. *Vanderbilt Law Review* 72:1497.
88. Crémer, J., D. Dinielli, P. Heidhues, G. Kimmelman, G. Monti, R. Podszun, M. Schnitzer, F. Scott Morton, and A. De Streel. 2023. Enforcing the Digital Markets Act: Institutional choices, compliance, and antitrust. *Journal of Antitrust Enforcement* 11 (3): 315–349.
89. Heath, M. K. 2018. What kind of (digital) citizen? A between-studies analysis of research and teaching for democracy. *The International Journal of Information and Learning Technology* 35 (5): 342–356.
90. Rosado, K. A. P., and K. J. R. Flores. 2025. Neurorights and teaching ethics: A new horizon for higher education. *NeuroData* 2:102–102.
91. Dame Adjin-Tettey, T. 2022. Combating fake news, disinformation, and misinformation: Experimental evidence for media literacy education. *Cogent arts & humanities* 9 (1): 2037229.
92. Morse, T. E. 2004. Ensuring equality of educational opportunity in the digital age. *Education and Urban Society* 36 (3): 266–279.
93. Priyanto, D. 2024. Education towards equality: A review of the contribution of distance learning technology in expanding access and promoting social equality. *Technology and Society Perspectives* 2 (1): 138–142.
94. Casilli, A. 2019. *En attendant les robots: Enquête sur le travail du clic*. Paris: De Seuil.
95. Ferraris, M. 2024. *Webfare: A manifesto for digital wellbeing*. Bonn: Transcript Verlag.
96. Van Deursen, A., and J. Van Dijk. 2011. Internet skills and the digital divide. *New Media & Society* 13 (6): 893–911.
97. Giacomini, G., & R. Paura 2023. The ideal of pluralism and the problem of online polarisation. Four scenarios and five proposals for the future. *Journal of Futures Studies,* 27(3).
98. Cinelli, M., G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini. 2021. The echo chamber effect on social media. *PNAS* 118 (9): e2023301118.
99. Van Bavel, J. J., S. Rathje, E. Harris, C. Robertson, and A. Sternisko. 2021. How social media shapes polarization. *Trends in Cognitive Sciences* 25 (11): 913–916.
100. Munroe, W. 2024. Echo chambers, polarization, and "Post-truth": In search of a connection. *Philosophical Psychology* 37 (8): 2647–2678.
101. Wollebæk, D., R. Karlsen, K. Steen-Johnsen, and B. Enjolras. 2019. Anger, fear, and echo chambers: The emotional basis for online behavior. *Social Media+ Society* 5 (2): 2056305119829859.
102. Acemoglu, D., A. Ozdaglar, and J. Siderius. 2024. A model of online misinformation. *Review of Economic Studies* 91 (6): 3117–3150.
103. Del Vicario, M., A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Sciences of the United States of America* 113 (3): 554–559.
104. Vaccari, C., and A. Chadwick. 2020. Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media+ Society* 6 (1): 2056305120903408.
105. Berlin, I. 1958. *Two concepts of liberty*. Oxford: Clarendon Press.