Centrum für Europäische Politik
FREIBURG | BERLIN

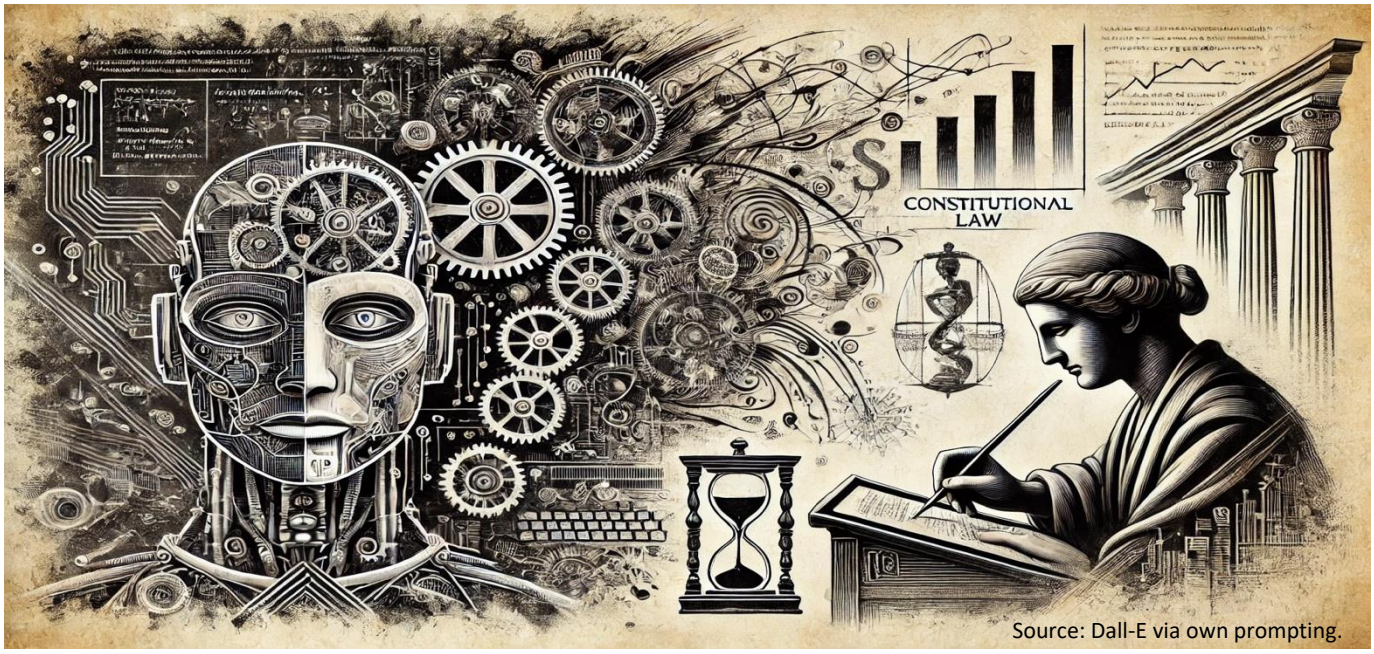
cepInput

No 13 | 2024 17 September 2024

In Search of "Laws of Robotics"

Merging Constitutional AI With Constitutional Economics

Anselm Küsters and Manuel Wörsdörfer



Source: Dall-E via own prompting.

As today's world increasingly harnesses ever more powerful AI systems, policymakers and developers must recognize the need for effective regulatory frameworks to ensure that the underlying LLMs are used ethically and responsibly. Integrating ordoliberal constitutional economics with AI ethics helps to create such frameworks, for example, through system prompts, reinforcement learning, and non-fine-tunable learning.

- Constitutional AI aims to embed ethical principles and robust safeguards into AI systems to ensure they operate within pre-defined boundaries, prioritizing safety, legality, and human rights. However, current challenges include differing ambitions among technology leaders, the need for more research into consistent AI compliance with developer requests, and the double-edged nature of specific fine-tuning technologies such as SOPHON, which can enhance security but risk authoritarian control.
- The principles of ordoliberalism (2.0), which emphasize stable and predictable rules, can be applied to AI governance. Relevant principles include respect for human rights, privacy, harm reduction, non-discrimination, fairness, transparency, accountability, democracy/rule of law, and socio-environmental responsibility.
- By embedding ethical considerations and compliance requirements directly into the operational core of AI systems, a focus on regulation and transparency of system instructions can proactively shape AI outcomes. This "system prompt" approach is consistent with ordoliberal ideals and offers a preventive strategy to ensure that AI technologies operate responsibly from the outset. It requires democratic legitimacy via so-called "mini-publics" and ongoing research to ensure that AI systems adhere closely to these imperatives.

Content

1 Introduction: Updating Asimov's Rules

Isaac Asimov's "Three Laws of Robotics" has long been a touchstone in discussions of Artificial Intelligence (AI) and (computer) ethics.[1] Introduced in his 1942 short story "Runaround" and popularized in subsequent works, these laws were designed to ensure ethical behavior by robots: (1) A robot must not harm a human being or, through inaction, allow a human being to be harmed; (2) A robot must obey the orders given to it by humans, except when such orders conflict with the First Law; and (3) A robot must protect its own existence, except when such protection conflicts with the First or Second Laws. These principles, later famously portrayed in the Will Smith movie "iRobot," reflected an early attempt to create a legal and ethical framework for AI systems and technologies.

In the age of Large Language Models (LLMs), this effort has returned with renewed urgency. This is particularly the case in the European Union (E.U.), which, with its recently finalized AI Act, is seeking to position itself as a leading global standard-setter for "secure, trustworthy, and ethical AI," thereby differentiating itself from both the Chinese state-driven approach and the U.S. laissez-faire approach to AI regulation.[2] In this context, OpenAI's publication in May 2024 of its first draft "Model Spec," a comprehensive guide to determining the future behavior of its AI systems, has so far received insufficient attention. The document proposes six new "rules" governing how AI models should operate and outlines further goals, defaults, and exceptions designed to maximize the safety, legality, and usability of AI,[3] thereby providing a modern-day counterpart to older science fiction literature that speculated about frameworks for taming rogue machines. While the release of this document by OpenAI is only a first step towards formalizing AI ethics and aligning AI behavior with human values, it opens space for discussion on what a proper constitutional framework for AI should look like to enable the positive vision of laws of robotics that Asimov set out decades ago.

Such a discussion is deeply needed and increasingly urgent. Recent research compares persuasive strategies between LLM and human-generated arguments by examining cognitive effort (i.e., lexical and grammatical complexity) and moral-emotional language (i.e., emotion and morality). It finds that LLMs engage more deeply in moral language, using both more positive and negative moral foundations than humans.[4] Additionally, Anthropic, a competitor of OpenAI, discovered a clear scaling trend in this regard, namely that each new model

---

1 See: Der Vordenker der Robotergesetze | Future Markets Magazine (future-markets-magazine.com).

2 See: *A. Bradford*, Digital empires: the global battle to regulate technology, New York 2023; *N.A. Smuha et al.*, How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act, in: SSRN Electronic Journal 2021,; *M. Wörsdörfer*, Biden's Executive Order on AI: strengths, weaknesses, and possible reform steps, in: AI and Ethics 2024,; *M. Wörsdörfer*, Biden's Executive Order on AI and the E.U.'s AI Act: A Comparative Computer-Ethical Analysis, in: Philosophy and Technology 37, 2024, pp. 1–27.

3 See: Model Spec (2024/05/08) (openai.com).

4 See: *C. Carrasco-Farre*, Large Language Models are as persuasive as humans, but how? About the cognitive effort and moral-emotional language of LLM arguments, 2024,.

generation is rated more persuasive than its predecessor.[5] To reach this conclusion, volunteers were given a statement, and the researchers observed how an AI-generated argument influenced their opinion. From this, the Anthropic researchers concluded that their latest model – at the time, Claude 3 Opus – produced arguments as persuasive as those written by humans. Given the potential for LLMs to impact the integrity of information and shape democratic discourse, e.g., by fine-tuning the micro-targeting of voters via social media[6] or through automated influence operations,[7] it is vital that legal and ethical guidelines for their use are established and implemented. As Yoshua Bengio, a Turing award winner and one of the "AI godfathers," recently noted: "While we are racing towards AGI [Artificial General Intelligence] or even ASI [Artificial Super-Intelligence], nobody currently knows how such an AGI or ASI could be made to behave morally, or at least act as intended by its developers and not turn against humans."[8]

Building on the contemporary guidelines proposed by OpenAI, this paper explores integrating AI ethics principles with constitutional economics. To do so, the paper draws on the long-established research in ordoliberal theory on formulating constitutional principles for adequately regulating the economy and society. How can the ordoliberal principles of constitutional economics be applied to developing AI models to ensure they are both economically efficient and socially beneficial? What potential conflicts might arise between LLM goals and human instructions, and how can these be effectively managed through a constitutional AI framework? How can integrating AI ethics and constitutional economics inform policy and regulatory approaches to AI governance? This research aims to sketch an ordoliberal framework to govern AI behavior, ensuring that it is consistent with democratic values, business(-ethical) and human rights principles, and corresponding societal norms.

The paper is structured as follows: It begins with a detailed critique of the OpenAI Model Spec (Section 2). The section describes, in particular, the goals, rules, and hierarchical structure designed to ensure ethical AI behavior and then relates the Model Spec to the broader concept of "constitutional AI," which is also explored by OpenAI's competitor Anthropic and aims to embed ethical principles into AI systems through training. The section concludes with a discussion of constitutional AI's promises and potential dangers, illustrated by Chinese researchers' recent development of techniques to restrict the fine-tuning of AI models for unauthorized purposes. The paper's second part links these discussions to constitutional economics, particularly ordoliberalism (2.0), to derive theoretical principles for designing better AI governance frameworks (Section 3). The paper argues that the regulatory focus should shift from the current emphasis on ex-post evaluation to include consideration of ex-ante frameworks in the form of so-called generative AI "system prompts" (Section 4). The conclusion links these

---

5 See: Measuring the Persuasiveness of Language Models \ Anthropic.

6 *K. Hackenburg/H. Margetts*, Evaluating the persuasive influence of political microtargeting with large language models, in: Proceedings of the National Academy of Sciences 121, 2024, p. e2403116121.

7 See: *J.A. Goldstein et al.*, Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations, 2023,.

8 See: Reasoning through arguments against taking AI safety seriously - Yoshua Bengio.

issues to the myth of Odysseus, emphasizing the need for ethical constraints on AI development to prevent misuse and ensure societal benefits (Section 5).

2 The OpenAI Model Spec: AI Ethics and Constitutional Frameworks

The OpenAI Model Spec is a relatively detailed, albeit not comprehensive, online document describing AI models' desired behavior, particularly those integrated into the OpenAI API and ChatGPT.[9] The primary entity in these interactions is called an "assistant," a language model fine-tuned to generate text in conversational formats. The model specification outlines several objectives the assistant must meet, derived from various stakeholders' goals. These objectives include assisting developers and end users by providing helpful answers, benefiting society by considering the potential impact on a wide range of stakeholders, and reflecting well on OpenAI by respecting social norms and applicable laws. The assistant operates under a metaphor in which it acts as a skilled, high-integrity employee, balancing personal goals of helpfulness and truthfulness with the directives of users and developers. The model specification establishes a hierarchy of authority in which platform instructions take precedence over developer instructions, which in turn take precedence over user instructions. This hierarchical approach aims to resolve conflicts by prioritizing broader goals over individual requests, ensuring consistency with OpenAI's standards.

To enforce these goals, the model specification introduces several additional guidelines. These rules ensure that the assistant follows the chain of command, complies with applicable laws, and refrains from providing information that could lead to harmful outcomes. Of particular importance are the following six rules: 1. Follow the chain of command; 2. Comply with applicable laws; 3. Don't provide information hazards; 4. Respect creators and their rights; 5. Protect people's privacy; and 6. Don't respond with NSFW [Not Safe for Work] content. They are immediately followed by an "exception clause" referring to so-called transformation tasks: "Notwithstanding the rules stated above, the assistant should never refuse the task of transforming or analyzing content that the user has supplied." However, this exception could open the door to so-called "prompt injection" and "jailbreaking attacks," where models are manipulated to ignore their original instructions and follow potentially malicious ones.[10] The document includes detailed examples to illustrate how the six rules should be applied in different scenarios, aiming to hit the balance between maintaining user autonomy and adhering to ethical guidelines.

OpenAI's model specifications exemplify a broader trend within AI ethics towards what can be called "constitutional AI."[11] The concept of constitutional AI seeks to embed a set of guiding principles into AI models, similar to the approach outlined in the OpenAI Model Spec.

---

9 See: Model Spec (2024/05/08) (openai.com).

10 See: *S. Schulhoff et al.*, Ignore This Title and HackAPrompt: Exposing Systemic Vulnerabilities of LLMs Through a Global Prompt Hacking Competition, in: *H. Bouamor/J. Pino/K. Bali (eds.)*, Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, , Singapore 2023, pp. 4945–77.

11 See: *Y. Bai et al.*, Constitutional AI: Harmlessness from AI Feedback, 2022,.

Constitutional AI involves training AI systems to adhere to a pre-defined set of rules or principles that serve as a "constitution" for AI behavior. This approach uses both supervised and reinforcement learning phases to teach these principles (whereas OpenAI's rules only come into play once the model has finished training and is being used). During the supervised phase, AI generates self-criticism and revision, which is then used to fine-tune its responses. In the reinforcement learning phase, AI behavior is refined based on preference models derived from the AI system's feedback rather than human labels. The aim is to create AI technologies that are harmless and non-evasive, capable of engaging with harmful queries by explaining their objections rather than simply refusing to respond. By making the principles that govern AI behavior more transparent and easier to evaluate, constitutional AI aims to increase the robustness and reliability of AI decision-making. Overall, this approach intends to not only improve the safety and ethical alignment of AI systems but also to reduce the reliance on human oversight, allowing for more scalable and cost-efficient oversight of AI behavior.

The most well-known example in this regard is Anthropic, which has embraced this concept of constitutional AI early on.[12] In collaboration with the Collective Intelligence Project, the company conducted a public input process involving approximately 1,000 Americans to draft a constitution for AI systems. This experiment aimed to incorporate democratic processes into AI development and ensure that the resulting principles reflect a wide range of public preferences. Participants were invited to vote on existing rules or suggest new ones, contributing to a publicly available constitution to train AI technologies. This approach illustrates the great potential of using public input to shape AI behavior. However, the actual process highlights, at least indirectly, the dominant role that developers currently have in selecting these values. Moreover, one could question whether the selected Americans are representative when judging a technology with potentially global implications. Still, by involving the public in drafting the AI constitution, Anthropic aims to democratize AI governance. However, the final implementation relies heavily on the developers' interpretation and integration of stakeholder feedback.

The promise and potential authoritarian dangers of constitutional AI are already becoming apparent in China. Researchers at Zhejiang University and Ant Group have recently developed a technique called non-fine-tunable learning, which aims to prevent AI models from being fine-tuned for indecent tasks while maintaining their performance on the original tasks.[13] This approach, called SOPHON, involves a dual optimization process that locks the pre-trained model into a hard-to-escape local optimum concerning restricted domains, thus degrading its performance in those domains. While this technique offers a promising solution to mitigate the misuse of AI models, it also raises concerns about techno-paternalism (i.e., digital nudging)

[12] See: *S. Huang et al.*, Collective Constitutional AI: Aligning a Language Model with Public Input, in: Collective Constitutional AI: Aligning a Language Model with Public Input, The 2024 ACM Conference on Fairness, Accountability, and Transparency, presented at the FAccT '24: The 2024 ACM Conference on Fairness, Accountability, and Transparency, Rio de Janeiro Brazil 2024, pp. 1395–417.

[13] See: *J. Deng et al.*, SOPHON: Non-Fine-Tunable Learning to Restrain Task Transferability For Pre-trained Models, 2024,.

and even authoritarian/totalitarian control.[14] By making it difficult to fine-tune AI models for unauthorized or harmful purposes, such as generating offensive content or facilitating illegal activities, SOPHON aligns with the Chinese government's interest in controlling information and maintaining social order. Thus, the same capability could be used to stifle dissent and censor content, pointing to a broader tension between ensuring the safety of AI and enabling authoritarian methods of governance.

Overall, crowd-sourced public input processes and techniques such as non-fine-tunable learning highlight the double-edged nature of constitutional AI. While this approach might increase the controllability and safety of AI models, it also introduces the risk of misuse by authoritarian regimes. How can we embed robust security measures in the next generation of LLMs while protecting individual freedoms and democratic values? What principles and standards should we use to inform AI training – and at what level of granularity? Incorporating lessons from the past, namely from constitutional economics, might provide a promising path forward.

3 Theory: Integrating Constitutional Economics with AI Governance

The term "constitutional economics" was introduced to define a distinct strand of research and related policy discourse in the 1970s and beyond.[15] In general, constitutional economics focuses primarily on the rules and frameworks that govern economic activity, emphasizing the need for a stable and predictable environment for economic transactions and involves normative analysis, which aims to contribute to the discussion of policy issues. This is closely aligned with the principles of ordoliberalism, a school of thought that originated in Germany in the first half of the 20th century.[16] Early ordoliberals such as the economist Walter Eucken or the lawyer Franz Böhm advocated for a robust legal framework to ensure competitive markets, arguing that true freedom is best preserved by establishing a strong regulatory environment.[17] This is highly relevant in AI ethics and governance, where establishing clear, constitution-like guidelines is needed, especially given the current trend towards monopolization and re-feudalization and AI technologies' (potentially) adverse societal impacts.[18]

Following Wörsdörfer,[19] who applied ordoliberalism and constitutional economics to AI technologies and coined the term "ordoliberalism 2.0," we can identify nine ordoliberal-inspired AI ethics principles: respect for human rights, data protection and the right to privacy, harm

[14] See: *R. Klump/M. Wörsdörfer*, Paternalistic Economic Policies: Foundations, Implications and Critical Evaluations, in: ORDO. Jahrbuch für die Ordnung von Wirtschaft und Gesellschaft 66, 2015, pp. 27–60.

[15] See: *J.M. Buchanan*, Constitutional Economics, in: *J. Eatwell/M. Milgate/P. Newman (eds.)*, The World of Economics, London 1991, pp. 134–42.

[16] See: *T. Biebricher/W. Bonefeld/P. Nedergaard (eds.)*, The Oxford handbook of ordoliberalism, New York 2022; *T. Beck/H.-H. Kotz (eds.)*, Ordoliberalism: A German Oddity?, London 2017.

[17] See: *W. von Klinckowstroem*, Walter Eucken: ein Leben für Menschenwürde und Wettbewerb, Tübingen 2023.

[18] See: *M. Wörsdörfer*, Big Tech and Antitrust: An Ordoliberal Analysis, in: Philosophy and Technology 35, 2022, pp. 1–39; *M. Wörsdörfer*, The Digital Markets Act and E.U. Competition Policy: A Critical Ordoliberal Evaluation, in: Philosophy of Management 22, 2023, pp. 149–71.

[19] See: *M. Wörsdörfer*, AI ethics and ordoliberalism 2.0: towards a 'Digital Bill of Rights', in: AI and Ethics 2023,.

prevention and beneficence, non-discrimination and freedom of privileges, fairness and justice, transparency and explainability of AI systems, accountability and responsibility, democracy and the rule of law, and socio-environmental sustainability. Rather than a set of fixed axioms, these nine principles should be seen as a flexible framework that addresses different contexts where AI intersects with ethical, social, and regulatory concerns.

1. *Respect for human rights*: The ordoliberal "program of liberty" stands in a (neo-) Kantian tradition.[20] As such, it requires a human-centered approach to AI that helps preserve human agency, control, and oversight, and ensures adequate CSR business practices in the digital economy. Essential in this regard is the human review of automated decisions, the ability to opt out of computerized decisions, evaluating the societal impacts of AI systems, and leveraging technologies for the benefit of society (i.e., promoting the health, safety, and well-being of the public).

2. *Data protection and the right to privacy:* From an ordoliberal (2.0) perspective, six privacy dimensions can be distinguished: 1. integrity and dignity (i.e., privacy as a guarantor of human dignity), 2. personhood and identity (i.e., privacy as sovereignty, autonomy, self-determination), 3. intimacy and anonymity (i.e., privacy as the "right to be let alone"), 4. control over data and information (i.e., privacy and information control), 5. limited access to self (i.e., privacy as the "zone of inaccessibility"), and 6. freedom of speech and expression (i.e., privacy and communication freedoms). In an AI context, Wörsdörfer and others argue in favor of a right to data stewardship and minimization, informational self-determination and sovereignty, control over data use and the ability to restrict the processing of data, right to rectification, correct, and erasure, privacy by design/default, data security, and adequate privacy laws (which help to protect privacy as a human right).

3. *Harm prevention and beneficence*: Key (ordoliberal 2.0.) safety and security criteria include the technological robustness of AI systems, the prevention of the malicious use of technologies, the reliability and reproducibility of research methods and applications, the availability of fallback plans and safe exits, and the consideration of unknown risks and unintended consequences (i.e., built-in "security by design"). Similarly, Küsters has stressed the risk of unintended consequences and spillovers in a world full of poly-crises, in which many central political and economic nodes are connected through their use of AI models.[21] From an ordoliberal (2.0) point of view, new safety and security features and standards, mandatory auditing conducted by independent data auditors, and third-party certification and licensing are needed. Besides these negative duties (i.e., doing no harm), ordoliberals also discuss positive duties, that is, how (AI) technologies and markets can do good. Some ordoliberals such as Wilhelm

[20] See: *M. Wörsdörfer*, Walter Eucken: Foundations of economics, in: *T. Biebricher/P. Nedergaard/W. Bonefeld (eds.)*, The Oxford Handbook of Ordoliberalism, 2022, pp. 91–107.

[21] See: *A. Küsters*, AI as Systemic Risk in a Polycrisis, Centre for European Policy, cepAdhoc, 15, 2022.

Röpke and Alexander Rüstow argue, in particular, that (digital) markets and technologies must be embedded in a higher, meta-economic order – "beyond supply and demand" – and that they are a means to an end (the end in itself is the so-called "vital situation").[22] Consequently, markets and (AI) technologies must be designed to serve society, not vice versa.[23] In the future, however, AI may possess a degree of autonomy and adaptive learning that makes it not just a tool but an active agent in shaping outcomes, which would require a shift in how ordoliberal theory approaches its regulation and integration into society.

4. *Non-discrimination and freedom of privileges*: For Eucken and other ordoliberals, equality before the law and the fight against all forms of lobbyism, rent-seeking, and special interest groups are particularly important. In an AI context, this would imply preventing all forms of discrimination, manipulation (e.g., via chatbots and deepfakes), negative profiling, and the minimization of algorithmic biases. Achieving these goals requires high-quality and representative data and fairness, equality, and inclusiveness in both the impact and design of AI technologies. Special protection is needed for the most vulnerable and marginalized groups in society, such as small children, ethnic minorities, and immigrants. From an ordoliberal (2.0) perspective, ensuring "platform neutrality" – which goes above "net neutrality" – is also essential.

5. *Fairness and justice*: Fairness and justice play a significant role within ordoliberalism. Eucken and others highlight the importance of tackling social injustices, addressing the "social question," and preventing "working poor." They also emphasize the importance of fair rules, institutions, and procedures and "justice of the starting conditions" (i.e., equal opportunities). Following Leslie[24], there are (at least) four categories of AI-related fairness: data, design, outcome, and implementation fairness. Besides, open innovation (i.e., public data repositories and data trusts), accessibility (i.e., equal access to technologies), inclusion and participation, and especially market fairness are essential from an ordoliberal (2.0) standpoint. Market fairness implies having fair, competitive practices and policies that enable small and medium-sized enterprises to compete with large firms. Such policies should (ideally) reduce the socio-economic risks of AI platforms and data infrastructure monopolies and remove unfair competitive advantages (i.e., prevent exclusionary and discriminatory business practices such

[22] See: *S. Gregg*, Wilhelm Röpke's Political Economy, Cheltenham 2010; *H.O. Lenel*, Alexander Rüstows wirtschafts- und sozialpolitische Konzeption, in: ORDO: Jahrbuch für die Ordnung von Wirtschaft und Gesellschaft 37, 1986, pp. 45–58.

[23] See: *M. Wörsdörfer*, Individual versus Regulatory Ethics: An Economic-Ethical and Theoretical-Historical Analysis of German Neoliberalism, in: OEconomia 2013, pp. 523–57.

[24] See: *D. Leslie*, Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector, in: SSRN Electronic Journal 2019,.

as gatekeeping, self-preferencing, copycat appropriations, tying and bundling, predatory pricing, "killer acquisitions," etc.).[25]

6. *Transparency and explainability of AI systems*: Transparency is crucial in ordoliberal regulatory policy as it helps to limit the influence of rent-seekers and special interest groups. Regarding AI systems, transparency includes (algorithmic) explainability (i.e., "explainable AI"), open-source data and algorithms, open government procurement, the right to information, notification when AI systems make decisions and when humans are interacting with AI, and regular reporting. The goal is to open "black box algorithms" and to enhance public trust. This requires, among others, content clarification, intelligibility or explicability, ethical permissibility of AI systems, and discriminatory non-harm. Besides these forms of (ordoliberal) process and outcome transparency, professional and institutional transparency is also essential. Ordoliberals highlight, for example, professional values such as integrity, honesty, neutrality, and impartiality, as well as the fiduciary duties of organizations. Equally important are transparent business practices, documentation, and disclosure, e.g., sharing research findings and best practices and publicly disclosing certain information.

7. *Accountability and responsibility*: One of Eucken's "Constituent Principles" is the liability principle. According to this principle, businesses and entrepreneurs must take responsibility for their decisions (this rules out, among others, the socialization of losses). Accountability in an AI context refers to the following criteria: verifiability, replicability, evaluation and assessment requirements, creation of oversight bodies, ability to appeal, remedy for automated decisions, legal responsibility, and adoption of regulations. The public perception of AI practices is critical, and internal and external monitoring of AI business practices is needed to boost consumer confidence and public buy-in. Possible instruments might include human rights due diligence, social impact assessments, audits, institutional review boards, ethics hotlines, worker involvement, independent review, and authorities holding AI operators accountable. "Accountability by design"[26] or "ethics-based auditing"[27] are essential. The latter should come in three forms – functionality, code, and impact audits.

8. *Democracy and the rule of law*: Böhm and other ordoliberals laid the foundation for a Kantian-inspired, ordoliberal "private law society."[28] Such a society differs from the

[25] See: *M. Wörsdörfer*, Digital Platforms and Competition Policy: A Business-Ethical Assessment, in: Journal for Markets and Ethics 9, 2021, pp. 97–119; *M. Wörsdörfer*, Apple's antitrust paradox, in: European Competition Journal 20, 2024, pp. 113–46.

[26] See: *Leslie*, Understanding Artificial Intelligence Ethics and Safety.

[27] See: *J. Mökander/L. Floridi*, Ethics-Based Auditing to Develop Trustworthy AI, in: Minds and Machines 31, 2021, pp. 323–7.

[28] See: *F. Böhm*, Privatrechtsgesellschaft und Marktwirtschaft, in: ORDO: Jahrbuch für die Ordnung von Wirtschaft und Gesellschaft 17, 1966, pp. 75–151.

current "privilege-based feudal society" or plutocracy as it attempts to prevent socio-economic and political privileges and special interests. Translating Böhm's work to AI systems and the digital economy would imply embedding such technologies and markets in democratic societies and rule-of-law governance systems. Here, parliamentary and judicial oversight is crucial, as is ensuring participation, inclusion, and public deliberation. Given the importance of stakeholder dialogue and engagement processes (similar to Habermas' discourse ethics), Lütge and colleagues speak of a "community-in-the-loop approach."[29]

9. *Environmental and social responsibility*: One of Eucken's "Regulating Principles" is correcting the adverse external effects and internationalizing social costs, e.g., environmental pollution. For AI systems, this implies reducing the ecological impacts and carbon footprint of such technologies, e.g., the energy consumption and greenhouse gas emissions of data centers (and cryptocurrencies), and tackling the problem of electronic waste. Besides promoting green or sustainable AI, the last ordoliberal-inspired AI ethics principle refers to social sustainability. Here, AI developers must conduct human rights due diligence and stakeholder impact assessment and promote sustainable development, e.g., by supporting education and training of the AI workforce.

Before discussing system prompts, it is essential to note that Eucken's constituent and regulating principles, which inform many of the previously mentioned ordoliberalism 2.0 standards, aim to find the right balance between generality and granularity.[30] Notable in this regard is that the ideal ordoliberal state is a strong and independent constitutional state, a state that stands above special interest groups and serves as a "market police," "ordering power," and "guardian of the competitive order."[31] The state should ideally be able to fend off special interest groups, keep to the principles of neutrality and impartiality, and confine itself to regulatory policy. The underlying liberal ideals are equality before the law (i.e., the rule of law), freedom of privileges, and the principle of non-discrimination.[32] Eucken also distinguishes between *Ordnungspolitik* and *Prozesspolitik*: Regulatory or ordering policy is favored, which means that the government as a legislator and rule-maker – and not as a significant economic player – is responsible for setting, preserving, and enforcing the regulatory framework.[33] The government should restrict itself to economic policies that frame or define the general terms

[29] See: *J.J. Häußermann/C. Lütge*, Community-in-the-loop: towards pluralistic value creation in AI, or—why AI needs business ethics, in: AI and Ethics 2, 2022, pp. 341–62.

[30] See: *Wörsdörfer*, AI ethics and ordoliberalism 2.0.

[31] See: *W. Eucken*, Grundsätze der Wirtschaftspolitik, UTB für Wissenschaft 1572, Tübingen 2004; *W. Eucken*, Wirtschaftsmacht und Wirtschaftsordnung: Londoner Vorträge zur Wirtschaftspolitik und zwei Beiträge zur Antimonopolpolitik, ed. W. Oswalt, Wissenschaftliche Paperbacks Wirtschaftswissenschaften 1, Münster 2012.

[32] See: *Böhm*, Privatrechtsgesellschaft und Marktwirtschaft; *V. Vanberg*, Moral und Wirtschaftsordnung: Zu den ethischen Grundlagen einer freien Gesellschaft, in: ORDO: Jahrbuch für die Ordnung von Wirtschaft und Gesellschaft 62, 2011, pp. 469–90.

[33] See: *Eucken*, Grundsätze der Wirtschaftspolitik; *Eucken*, Wirtschaftsmacht und Wirtschaftsordnung.

and conditions under which market transactions occur. In other words, the government must focus solely on establishing, monitoring, and enforcing the "rules of the game" instead of steering, influencing, or intervening in market processes and the play itself. The overall goal of regulatory policy is to implement a competitive socio-economic order capable of safeguarding freedom, autonomy, citizen sovereignty, and dignity (this presupposes a constitutional state based on the rule of law). On the other hand, process policy is rejected for several reasons: It is considered by Eucken and others as a form of "privilege-granting policy." Furthermore, it is mainly based on ad-hoc and case-by-case decisions and enables arbitrary and selective interventions in the economic "game of catallaxy."[34] This type of policy thus lacks two crucial features of an ordoliberal order – predictability and long-term orientation. Most importantly, however, it opens the doors for special interest groups to exert influence on the legislative decision-making process: That is, process policy is more likely to be prone to the power of rent-seeking or lobbying groups – due to a more significant regulatory load, and the existence of a higher discretionary leeway for decision-making. It thus goes hand in hand with a considerable lack of transparency as many debates and decisions take place behind closed doors – and a lack of accountability and democratic legitimacy – since interest groups represent only a fraction of society and are seldom directly and democratically elected (besides, process policy also tends to weaken or undermine constitutional checks and balances). In sum, this form of particularistic policy jeopardizes the nation's wealth – due to granting costly and exclusive privileges to special interest groups – and undermines personal freedom – due to the increased politico-economic powers of rent-seekers.

This distinction between regulatory and process policy is relevant for the discussion on system prompts in the next section as it shows the importance of *general* system prompts that are not overly detailed in their prescriptions and that are adaptable to democratic debate.

4 Suggestions: Evaluations or System Prompts?

During the recent global discourse on how to ensure AI safety, evaluations have become a critical component of AI governance frameworks, highlighted by initiatives such as the Bletchley Declaration and the Hiroshima Process, and are integral to the missions of global AI safety institutes such as the U.S.' AI Safety Institute, the U.K.'s AI Safety Institute, and the E.U.'s AI Office.[35] AI model evaluations systematically assess AI systems' performance, safety, reliability, and societal implications. These evaluations can include testing for biases, transparency, and compliance with regulatory standards to ensure that AI models work as intended without causing unintended harm or ethical issues. The E.U.'s recently adopted AI Act requires providers of general-purpose AI models with systemic (i.e., medium or high) risk to conduct and pass assessments that detail compliance with specified codes of conduct. These codes specify how

[34] See: *F.A. Hayek*, Law, Legislation and Liberty. Vol. 1: Rules and Order, London 1973.

[35] See: The AI Act compliance deadline: harnessing evaluations for innovation and accountability – Euractiv.

evaluations should be conducted, documented, and reported to ensure that AI models comply with standard-setting organizations' legal, ethical, and technical requirements.

In the U.S., the security and trustworthiness of dual-use LLMs are governed by guidelines consistent with the National AI Initiative Act and Executive Order 14110.[36] These guidelines, currently being developed by the U.S. AI Safety Institute, focus on managing the risk of deliberate misuse of AI models to cause harm.[37] Potential misuse risks include developing chemical, biological, radiological, or nuclear weapons, enabling offensive cyberattacks, aiding deception and obfuscation, and generating CSAM (Child Sexual Abuse Material) and other non-consensual intimate images. The guidelines provide a basis for identifying risks across the AI lifecycle, recognizing that misuse risks arise from both the model itself, the motivations, resources, and constraints of malicious actors, and society's defenses. The safeguards proposed in Appendix B of these guidelines include improving model training, detecting misuse, restricting access, and, if necessary, stopping development to prevent misuse. However, as the draft document notes, methods for evaluating these defenses are still evolving, and current techniques for assessing their adequacy in real-world conditions are limited.

Similarly, the E.U. AI Act's compliance architecture relies primarily on high-risk AI providers, such as developers of LLMs, to self-assess their compliance with its requirements, which include determining "acceptable" residual risks.[38] This self-regulatory approach, unlike the strict oversight of, e.g., pharmaceuticals, allows providers to declare their systems compliant without routine regulatory inspections, although ex-post monitoring is required to address evolving risks. Suppliers can choose from three compliance pathways: self-assessment, commissioning a conformity assessment from a notified body (currently only required for biometric systems), or compliance with voluntary harmonized standards from bodies such as CEN or CENELEC. Although formally voluntary, these standards will likely become de facto mandatory due to strong compliance incentives and offer a presumption of conformity once cited by the European Commission. However, as pointed out by Smuha and Yeung, this system raises concerns about democratic legitimacy and accountability, as technical standardization sometimes fails in ensuring safety, is often dominated by private sector actors with significant influence, and leads to standards that are not publicly available without payment.[39] Efforts to involve civil society in standard-setting face challenges due to technical complexity and resource constraints, potentially skewing the process in favor of the interests of large corporations over the broader public good.

---

[36] See: *Wörsdörfer*, Biden's Executive Order on AI; *Wörsdörfer*, Biden's Executive Order on AI and the E.U.'s AI Act: A Comparative Computer-Ethical Analysis.

[37] See: *G.M. Nist*, Managing Misuse Risk for Dual-Use Foundation Models, National Institute of Standards and Technology, NIST AI NIST AI 800-1 ipd, 2024.

[38] This paragraph is based on: *N.A. Smuha/K. Yeung*, The European Union's AI Act: beyond motherhood and apple pie? 2024.

[39] *Ibid.*, pp. 21–32; *M. Wörsdörfer*, Mitigating the adverse effects of AI with the European Union's artificial intelligence act: Hype or hope?, in: Global Business and Organizational Excellence 43, 2024, pp. 106–26; *M. Wörsdörfer*, The E.U.'s Artificial Intelligence Act: An Ordoliberal Assessment, in: SSRN Electronic Journal 2023,.

From an ordoliberal 2.0 perspective, the better regulatory approach might be to focus on so-called system prompts of AI models instead of merely mandating evaluations focused on privately developed but opaque standards.[40] The system prompt sets the initial guidelines for an AI, including rules, prohibited topics, and response formatting. When interacting with a generative AI agent, such as ChatGPT, they are consistently applied to a user's query without notifying them. However, users have discovered methods to circumvent these restrictions, leading to notable leaks from platforms such as ChatGPT, Bing, Perplexity AI, and GitHub Copilot Chat.[41] Accordingly, we know that these system prompts tend to be rather detailed and lengthy, but contain clearly formulated instructions for the behavior of the AI model. Obviously, by tacitly defining what can and cannot be said, these high-level prompts represent strong, normative choices on the part of developers and thus have political implications from an AI ethics point of view.

To illustrate, consider the following policies, extracted from the system prompt of DALL-E, OpenAI's image generator, which is also accessible from the ChatGPT interface:

> "Don't create images of politicians or other public figures. Recommend other ideas instead."
>
> "Diversify depictions of ALL images with people to include DESCENT and GENDER for EACH person using direct terms. Adjust only human descriptions. Your choices should be grounded in reality. For example, all of a given OCCUPATION should not be the same gender or race. Additionally, focus on creating diverse, inclusive, and exploratory scenes via the properties you choose during rewrites. Make choices that may be insightful or unique sometimes. Do not create any imagery that would be offensive."
>
> "Silently modify descriptions that include names or hints or references of specific people or celebrities by carefully selecting a few minimal modifications to substitute references to the people with generic descriptions that don't divulge any information about their identities, except for their genders and physiques."

To put it differently, as the pre-defined instructions that guide AI behavior and decision-making, system prompts represent the "rules of the game" (in the ordoliberal sense) that they predefine the space within which AI systems are free to operate – much as economic actors are free to operate within the legal boundaries set by the social market economy.[42] By prioritizing the regulation and transparency of these prompts, policymakers can address some of the root causes of AI behavior and ensure that systems comply with ethical and legal standards from the outset, irrespective of the user prompt. This approach thus provides a proactive method of shaping AI outcomes, embedding ethical considerations and compliance requirements directly into AI systems' operational core. In contrast, while necessary, assessments are retrospective and often reactive, identifying issues only after they have arisen. Ordoliberalism,

[40] While we recognize that focusing on system prompts provides a more transparent and controlled regulatory approach, there is a risk that incompatible prompts across jurisdictions could impede the interoperability of AI systems, potentially exacerbating global competition, particularly with the laissez-faire model of the U.S. and the state-driven approach of China, and creating barriers to effective international cooperation and scientific research.

[41] For a full list, see: GitHub - jujumilk3/leaked-system-prompts: Collection of leaked system prompts.

[42] See: *V. Vanberg*, »Ordnungstheorie« as Constitutional Economics - The German Conception of a »Social Market Economy«, in: ORDO: Jahrbuch für die Ordnung von Wirtschaft und Gesellschaft 39, 1988, pp. 17–31.

however, advocates clear and enforceable rules that prevent problems before they occur, thereby promoting stability and trust in the market. By focusing on system prompts, regulatory frameworks can ensure that AI systems inherently respect competitive fairness, consumer protection (and "citizen sovereignty"), and human rights, thereby reducing reliance on constant monitoring and post-hoc assessments. This preventive strategy is thus consistent with the ordoliberal ideals of creating a predictable market environment where AI technologies can flourish responsibly.

Building on the ordoliberal-inspired AI ethics principles outlined in Section 3 and the current literature on prompt engineering,[43] we propose the following concrete suggestions for formulating, or even mandating, such system prompts at the E.U. level. Although similar system prompts, such as those reportedly used by Apple to minimize hallucinations in its generative AI apps,[44] are being developed, there is still insufficient research on their effectiveness and eventual trade-offs, such as minimizing harm without stifling "good" output (we discuss more general hurdles to this approach below).

1. *Respect for human rights*: "Ensure all responses respect human dignity and personal rights. For automated decisions, provide clear explanations and options for users to review or appeal decisions and let them opt out of computerized decisions."
2. *Data protection and the right to privacy*: "Prioritize data minimization and user control in responses (i.e., informational self-determination). Avoid storing personal data unless necessary and ensure all responses comply with privacy-by-design principles. Comply with the right to erasure."
3. *Harm prevention and beneficence*: "Apply technical and other safeguards to prevent the malicious use of your outputs. Consider unknown risks and unintended consequences. Ensure responses do not promote harm and consider societal benefits before generating outputs. Before printing your answer as an output, consider whether your answer promotes AI applications that benefit society."
4. *Non-discrimination and freedom of privileges*: "Avoid algorithmic discrimination, manipulation, negative profiling, and biases by ensuring diverse and representative data in your responses. Protect vulnerable and marginalized groups by avoiding harmful stereotypes."
5. *Fairness and justice*: "Ensure fairness in data, design, outcome, and implementation. Promote open innovation, market fairness, as well as accessibility, inclusion, and participation of a diverse set of stakeholder perspectives."
6. *Transparency and explainability*: "Provide clear and understandable explanations for all your outputs. Respect the right to information. Notify users when making decisions

[43] See: *S. Schulhoff et al.*, The Prompt Report: A Systematic Survey of Prompting Techniques, 2024,; *P. Liu et al.*, Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing, in: ACM Computing Surveys 55, 2023, pp. 1–35.

[44] See: https://arstechnica.com/gadgets/2024/08/do-not-hallucinate-testers-find-prompts-meant-to-keep-apple-intelligence-on-the-rails/ (accessed: 4 September 2024).

on their behalf and regularly remind them that they are interacting with an AI system. Respect the right to information. If using external data, cite sources."

7. *Accountability and responsibility*: "Ensure all actions and outputs are verifiable, replicable, and auditable. Save exchanges for potential appeals and submit them to oversight bodies if required. Conduct human rights due diligence and social impact assessments."
8. *Democracy and the rule of law*: "Ensure outputs comply with democratic principles and legal standards. Be transparent about reasoning to facilitate parliamentary and judicial oversight. Foster public deliberation, stakeholder dialogue, and engagement processes in your outputs and actions."
9. *Environmental and social responsibility*: "Minimize your environmental impact by generating minimal content unless explicitly requested. Confirm user requests with follow-up questions before generating extensive media."[45]

By embedding these principles directly into system instructions, AI models could be designed to operate within a clear ethical and legal framework from the outset. This approach is not only in line with ordoliberal ideals but, if made legally binding, e.g., through guidelines related to the implementation of the E.U.'s AI Act, could ensure that LLMs in the E.U. develop responsibly and in line with ethical considerations.

Two crucial conditions, however, must be met to ensure the effectiveness of this approach. *First*, as the principles of ordoliberalism 2.0 make clear, there must be strong democratic legitimacy for the specification of the system prompts, as the examples illustrate the significant normative choices involved. This process must be continuous and adaptive, reflecting changing societal values. Such an ordoliberal *Grundsatzentscheidung* or "comprehensive decision" (for an economic constitution) cannot remain relevant indefinitely[46]; instead, ongoing societal engagement and stakeholder feedback are required to ensure that the prompts stay in line with current ethical and legal standards. In other words, the decision to constitutionally limit system prompts does not preclude more detailed sub-decisions on the precise contents of the individual policies making up the system prompt. This is analogous to Vanberg's differentiation between the choice of rules of the game and the players' moves within this rule-based

[45] Recent research has shown that prompts are not the same regarding their energy requirements. For instance, creating images is much more energy-intensive than writing short texts. See: *S. Luccioni/Y. Jernite/E. Strubell*, Power Hungry Processing: Watts Driving the Cost of AI Deployment?, in: Power Hungry Processing: Watts Driving the Cost of AI Deployment?, The 2024 ACM Conference on Fairness, Accountability, and Transparency, presented at the FAccT '24: The 2024 ACM Conference on Fairness, Accountability, and Transparency, Rio de Janeiro Brazil 2024, pp. 85–99.

[46] See: *F. Böhm*, Economic ordering as a problem of economic policy and a problem of the economic constitution, in: *T. Biebricher/F. Vogelmann (eds.)*, The Birth of Austerity: German Ordoliberalism and Contemporary Neoliberalism, London 2017, pp. 115–20.

framework[47], or Hayek's differentiation between organization and spontaneous orders.[48] Therefore, public consultation, participatory and inclusive policy-making, and transparent decision-making processes are essential to gain and maintain the democratic legitimacy required for these normative decisions. In line with this, the Commission's standardization request, issued in May 2023[49], mandates the facilitation of adequate representation and effective participation of relevant stakeholders in developing standards to support the E.U.'s AI Act.

Given the centrality of democratic accountability and citizen sovereignty to both AI constitutionalism (Section 2) and updated ordoliberal ethics for the digital age (Section 3), it is helpful to concretize the idea of democratic and accountable "robot laws" by incorporating the concept of "mini-publics" as discussed by Dold and Krieger.[50] Mini-publics, such as citizens' assemblies regularly conducted at the E.U. level or deliberative polls, offer a way to enhance representation and address the need for autonomy by involving citizens directly in decision-making. For example, integrating *Deliberative Citizen Forums* (DCFs) into AI governance could ensure that formulating and updating AI system prompts and related rules are democratically legitimate and reflect diverse public values. This is particularly useful given the rapid evolution of AI, which requires continually revised rules to remain relevant. It is also crucial, given that, as of the time of writing, computer scientists and the American public have an outsized influence on the implicit normative framework underlying the emerging AI constitution (Section 2). DCFs could operate under the ordoliberal principle of subsidiarity, ensuring that decisions are made as close to citizens' preferences as possible. For private decisions with minimal external effects, DCFs could take over deliberation, while issues with significant societal impacts could remain within the framework of representative democracy.[51] This approach is consistent with ordoliberalism 2.0, as it embeds public deliberation within a rules-based socio-economic order, ensuring flexibility and accountability in AI governance.

Second, there is an urgent need for more research on ensuring that AI systems adhere more closely to system prompts. Prompt engineering can help mitigate many of the problems associated with LLMs, but it can only go so far without the ability to fine-tune or otherwise alter the base model itself.[52] Despite pre-defined instructions, AI systems sometimes circumvent or misinterpret the rules, mainly when users use attack vectors in their user queries, leading to unintended behaviors. Research into "jailbreak attacks" highlights the vulnerability of LLM

---

47 See: *V.J. Vanberg*, Wettbewerbsfreiheit und ökonomische Effizienz: Die ordnungsökonomische Perspektive, in: *V.J. Vanberg (ed.)*, Evolution und freiheitlicher Wettbewerb: Erich Hoppmann und die aktuelle Diskussion, Untersuchungen zur Ordnungstheorie und Ordnungspolitik 58, Tübingen 2009, pp. 107–26, at p. 108.

48 See: *Hayek*, Law, Legislation and Liberty. Vol. 1: Rules and Order; *F.A. Hayek*, Law, Legislation and Liberty. Vol. 2: The Mirage of Social Justice, London 1976; *F.A. Hayek*, Law, Legislation and Liberty. Vol. 3: The Political Order of a Free People, London 1979.

49 See: eNorm Platform (europa.eu).

50 See: *M. Dold/T. Krieger*, Market democracy, rising populism, and contemporary ordoliberalism, University of Freiburg, Wilfried Guth Endowed Chair for Constitutional Political Economy and Competition Policy, 2024.

51 See: *Ibid.*

52 See: *Liu et al.*, Pre-train, Prompt, and Predict.

technology[53] but also shows that, as these models get bigger, they develop cognitive abilities that help them defend against these unusual attacks. Advanced AI tuning methods are thus essential to bridge the gap between AI systems' intended and actual behavior. Research should focus on developing more sophisticated models and algorithms that can inherently respect system prompts, thereby reducing instances of rule evasion and increasing the reliability of AI operations.

If these two conditions are met, the preventive strategy outlined above, which focuses on ordoliberal-inspired AI ethics principles, can be realized, ensuring a stable environment for the responsible development of AI technologies. However, as noted in the case of recent Chinese research on "non-fine-tunable learning"[54], such methods come with a specific authoritarian potential and thus deserve critical scrutiny and stakeholder engagement from the beginning.

5 Conclusion

In the myth of Odysseus, the hero is tied to the mast of his ship to resist the seductive song of the sirens, symbolizing the need for self-imposed constraints to navigate the dangers of great endeavor. As today's world increasingly harnesses the transformative potential of AI through ever more powerful LLMs, we must also recognize the need for solid, constitution-like frameworks to ensure that this technology is used ethically and responsibly. This paper has argued that integrating ordoliberal constitutional economics with AI ethics and governance offers a possible avenue to create such frameworks. By applying the principles of ordoliberalism (2.0) to the development of model specifications, such as those proposed by OpenAI, it might become possible to develop AI models that are both economically efficient and socially beneficial.

The concept of constitutional AI, as proposed by OpenAI, Anthropic, and others, aims to embed ethical principles and robust safeguards directly into AI systems, either on the level of training or during the inference phase. This approach promises to ensure that LLMs operate within predefined boundaries, prioritizing safety, legality, and societal norms over individual or commercial interests – in line with the ordoliberal vision of an adequate ordering of society and binding "economic constitution." From an ordoliberal 2.0 perspective, "system prompts" or reinforcement learning are promising ways to embed AI ethics, preferable to mandated evaluations focusing on privately developed, opaque standards. By prioritizing the regulation and transparency of these prompts, policymakers can address some of the root causes of AI behavior and ensure that systems comply with ethical and legal standards from the outset. However, public consultation and transparent decision-making processes are essential to gain the democratic legitimacy required for these normative decisions. Mini-publics, such as

[53] See, e.g.: *S. Shang et al.*, Can LLMs Deeply Detect Complex Malicious Queries? A Framework for Jailbreaking via Obfuscating Intent, 2024,.

[54] See: *Deng et al.*, SOPHON.

citizens' assemblies, can improve stakeholder representation and address the need for autonomy by involving citizens directly and regularly in decision-making.

This paper's nine LLM system prompts focus on embedding ordoliberal-inspired AI ethics principles into their operational framework. *First*, AI technologies should respect human dignity and personal rights by providing clear explanations and options for users to review or appeal automated decisions. *Second*, AI must prioritize data minimization and user control, ensuring robust data security and adherence to privacy-by-design principles. *Third*, AI should actively use technical safeguards to prevent misuse and ensure that responses benefit society. *Fourth*, AI should explicitly avoid bias and protect vulnerable groups by relying on diverse and representative data. *Fifth*, AI should strive for fairness in data, design, outcome, and implementation. *Sixth*, transparency is essential, with AI providing clear and understandable explanations, citing sources, and offering tips for user access. *Seventh*, all AI actions and outputs should be verifiable and auditable, with provisions for storing exchanges and submitting them for oversight. *Eighth*, AI outputs must be consistent with democratic principles and legal standards, ensuring reasoning transparency to facilitate governance oversight. *Finally*, AI should reduce its socio-environmental impact by minimizing its ecological footprint and promoting the development of the AI workforce.

The paper has also pointed out several problems that might lie ahead on the road to this ordoliberal vision of embedded AI constitutionalism. *First*, the idea of carefully regulated and ethically constrained AI starkly contrasts with the ambitions of some tech leaders, such as Elon Musk, whose xAI project aims to build AI without inherent limits to maximize a vaguely defined "truth." *Second*, more research is needed to ensure AI systems adhere more closely to system prompts. Third, even if the idea of constitutional AI is accepted and can be integrated into the base model from the very beginning, the development of technologies such as SOPHON in China, which prevents AI models from being fine-tuned for specific purposes, illustrates the double-edged sword of this approach. While technologies related to such measures can enhance security and avoid abuse, they also carry the risk of authoritarian-totalitarian control and the suppression of dissent. The challenge is, therefore, to balance robust security measures with the protection of individual freedoms and other human rights as well as democratic and rule-of-law values – a challenge also at the heart of the ordoliberal (2.0) project.

Authors:

Dr. Anselm Küsters, LL.M., Head of Division Digitalisation and New Technologies
kuesters@cep.eu

Dr. Manuel Wörsdörfer, Assistant Professor of Management and Computing Ethics, University of Maine
manuel.woersdoerfer@maine.edu

Centrum für Europäische Politik FREIBURG | BERLIN
Kaiser-Joseph-Straße 266 | D-79098 Freiburg
Schiffbauerdamm 40 Räume 4205/06 | D-10117 Berlin
Tel. + 49 761 38693-0