Bylieva, D. S. and A. Nordmann. 2025. "Ontolytic Effects of AI: Widening the Framework for Responsible Research and Innovation" [in English]. *Filosofiya. Zhurnal Vysshey shkoly ekonomiki [Philosophy. Journal of the Higher School of Economics]* 9 (4), 84–104.

Daria Bylieva and Alfred Nordmann*

Ontolytic Effects of AI**

Widening the Framework for Responsible Research and Innovation

Submitted: Sept. 13, 2025. **Reviewed:** Oct. 10, 2025. **Accepted:** Oct. 18, 2025.

Abstract: This article examines how artificial intelligence (AI) disrupts the conceptual foundations of Responsible Research and Innovation (RRI). We argue that AI's ontolytic capacity — its ability to decompose and reconfigure established categories of authorship, cultural representation, and governance — renders conventional RRI frameworks inadequate. Focusing on generative AI's role in producing and distorting cultural identities, we demonstrate how the technology functions as both a mirror and an agent of societal values. The emergence of "national AI" systems further complicates this landscape by embedding particular cultural and ideological commitments into technical infrastructures. We contend that RRI must evolve beyond its Western-centric origins. It can no longer limit itself to managing the impacts of a neutral technology but must now navigate a landscape where AI is simultaneously a subject of governance, an agent in the governance process, and a battleground for global cultural and political influence. The future of RRI lies in its ability to address this tripartite challenge, fostering mechanisms for genuine epistemic inclusion in a world where the very concept of responsibility is being digitally deconstructed.

Keywords: Ontolytic Effect, Responsible Research and Innovation, Artificial Intelligence (AI), Large Language Models, Bias and Correctness, LLMs, Technological Sovereignty.

DOI: 10.17323/2587-8719-2025-4-84-104.

Two Problems of Dissociation

Many critiques of technological development take as their starting point a historical and socio-political analysis that argues that technologies are more than the sum of their technical functions and transport implicit cultural values. This is often discussed as a process of globalization that promotes Western hegemonialism. Advancing on the rails of capitalism,

*Daria Bylieva, PhD in Political Science; Associated Professor at Peter the Great St. Petersburg Polytechnic University (Saint Petersburg, Russia), bylieva_ds@spbstu.ru, ORCID: 0000–0002–7956–4647; Alfred Nordmann, PhD in Philosophy; Professor at Darmstadt Technical University (Darmstadt, Germany), nordmann@phil.tu-darmstadt.de, ORCID: 0000–0002–2173–4084.

the West no longer colonizes territories but colonizes hearts and minds. It elicits a consumerist mentality that valorizes individualization through private purchasing choices. Quality of life becomes defined, to a great extent, by what kind of technical infrastructure one can afford to create for oneself.

Arguably, certain applications of generative AI and the advance of digital technologies, more generally, complicate this narrative of Westernization through technological modernization: "AI takes shape in the multipolar world of TikTok and X, ChatGPT and DeepSeek, Apple and Huawei, WeChat and WhatsApp, Tesla and BYD. This is a world of social media platforms under suspicion, blocked here, allowed there, a world where chip manufacturing sometimes includes malicious capabilities, a world of export and import controls, of different privacy policies and ecological costs for browsers and their search engines" (Wang et al., eds., 2026). Does this new balancing of technological powers in a multipolar world signify, however, that the development of digital technologies is becoming dissociated from Western logics and values? The answer to this question is all but obvious. On first glance, there are reasons for being skeptical: In terms of functionality and the aesthetics of interface design, most of these technologies in East and West are near clones of each other, and more often than not they serve to promote purchasing behavior. Even the ethics discourse tends to revolve around some of the same issues such as the questions of "privacy" — which is not even a Western but specifically a U. S.-American value.[1] Perhaps, if there were an incipient debate about the ethical program of "value-alignment" — as opposed to value-sensitive design — it might be construed as a confrontation of Western and non-Western points of view (ibid.). This requires a close look at who argues what for whom. Indeed, quite generally: If one wants to look for the seeds of differentiation and dissociation, perhaps rebellion, these are most likely to be found behind the scenes in how certain technologies play out in different cultural contexts. To look behind the scenes, therefore, is to look for the disruptiveness of generative AI, even and especially when at first glance it appears culturally homogeneous. The very seamlessness of

[1]Enlightenment Europe speaks of a right to self-determination and thus of a protective sphere for that right. In the age of communication technologies, this was extended to include "informational self-determination," that is, a sovereign power over the keeping and sharing of information about oneself. To be sure, this prepared the ground for the passively construed "right to privacy" but should not be equated with it. While dissident subjects of authoritarian states might evidently benefit from such rights, it is not at all obvious why societies that are built on social solidarity and personal sacrifice for the common good should be interested in "privacy" except to the extent that it is rooted in the pyschology of personal shame.

the associations forged within Large Language Models (LLMs) may hold the seed for their destruction.

A second problem of dissociation with respect to the governance of technological development. To be sure, there are industrial policies and state-sponsored technical developments — such as space programs — that channel national ambitions. In addition, there is in some form or another always the concern about whether the diffusion and integration of technologies serve to express general ideas of social progress and human advancement. The most distinctive example of this was the idea in Soviet times of progressively forging a new human and a new humanity for the machine age. A very distant cousin of this endeavor can be found in the notion of "Responsible Research and Innovation" (RRI). It is very much a product of the European Union (EU) not only for internal purposes but, with its universalist aspirations, also for the explicit export of European values. Originating in the contest of European and US-American schemes for internationalizing the societal discourse about nanotechnology and the so-called converging technologies, the EU believed it had the more attractive offer to make — more open and procedural, more inclusive and humble (Felt, 2007; Jasanoff, 2002; 2005; Nordmann, 2009). In the European context of STS-inspired policies, RRI fits into a long succession of schemes to build trust and ensure the integration of research and innovation with the values and goals of European societies. Alongside the parliamentary institution of "Technology Assessment" came "Science Cafes" and discussions of professional and lay expertise (Collins & Evans, 2002; Wilsdon et al., 2005). So-called "geistes- und sozialwissenschaftliche Begleitforschung" performed investigations "alongside" or in "parallel" to scientific and engineering research, bringing to bear their own type of scientific authority to what was known as the "Ethical, Social, and Legal Aspects" or "Implications" (ELSA and ELSI) of emerging technologies.

After the so-called "GMO disaster" and public rejection particularly of genetically modified crops, the focus turned to the "responsible development" of emerging technologies, culminating in "NEST-ethics" and a "Code of Conduct for Responsible Nanosciences and Nanotechnologies Research" which was successfully marketed to other countries (Brazil, South Africa, Japan, South Korea) while failing to gain the requisite support of European member states, some of which considered it too radical. If all these schemes had technologies in the driver's seat, seeking ways to modulate their trajectory and rendering them compatible with the values and concerns of citizens (Fisher et al., 2006), RRI would take a radical turn, assigning to science and engineering a subsidiary role. Distrustful of the internal dynamics, reward

systems, publication economics, redundancies, and developmental logic of academic and commercial research (Von Schomberg, 2019), RRI prepared the ground for more recent European policy concepts such as the "3 Os" of "open science" (Open Innovation, Open Science, Open to the World..., 2016). Positing an alternative to research-business as usual, RRI thus shifts the focus to technologies that are emerging from the labs, by putting challenges and problems first, inviting researchers to contribute to their solution within the framework of European values (Von Schomberg, 2015).[2] The challenges include resource depletion, environmental degradation, climate change, but also social inequities, smart cities, and the like. The "European values" are those included in the Lisbon Treaty from December 2007 which went into effect in 2009 as the political constitution of the EU:

> Article 3
>
> 1. The Union's aim is to promote peace, its values and the well-being of its peoples.
>
> 2. The Union shall offer its citizens an area of freedom, security and justice without internal frontiers [...].
>
> 3. The Union shall [...] work for the sustainable development of Europe based on balanced economic growth and price stability, a highly competitive social market economy, aiming at full employment and social progress, and a high level of protection and improvement of the quality of the environment. It shall promote scientific and technological advance.
>
> It shall combat social exclusion and discrimination, and shall promote social justice and protection, equality between women and men, solidarity between generations and protection of the rights of the child (Treaty on European Union..., 2008).

Reading this article, it is easy to see why Europeans are generally convinced of its universal appeal: If these are "Western values," why would anyone prioritize "non-Western" values instead? If this question is accompanied by a lack of imagination for what these other values might be, it serves as evidence for the arrogance of the West. But if there were perhaps such arrogance on one side, it becomes quite difficult on the other side to dissociate a productive notion of "Responsible Research and Innovation" from European values. This second problem of dissociation is even harder than the first one, since RRI was designed to be a vehicle for them. And

[2]Here is one way in which AI is disruptive of the RRI-philosophy: It appears to be technology-driven, indeed, moving forward along its trajectory by an internal logic with utter disregard for societal needs. But is this, perhaps, only one of the myths about AI: Does it not move forward through the conjunction of many very specific demands for surveillance, automation, globalized consumerism, or by commensuration and the concentration of power?

yet, the very proceduralism of Western notions of democracy also makes them a vessel that accommodates many cultural specificities, including anti-capitalist notions. Vehicle for Europeanness and vessel for cultural difference — RRI appears to be both and this reveals at least one Western value that does not travel well: RRI is expressive of a purely formal social framework that demands tolerance, if not indifference towards different notions of the good life as if these notions could coexist without contradiction. Through the dogmatic pluralization of cultural identities, cultural identity is simultaneously affirmed and denied.[3] But if one were to abandon this dogmatic requirement of tolerance, inclusivity, or multi-perspectivalism, can we still speak of and draw upon RRI at all? Here again, generative AI appears to hold a key. It exposes the tension by appearing to fuse cultural horizons and by manifesting the impossibility of doing so.

ONTOLYSIS

Generative AI and LLMs may pose some new philosophical problems but more importantly they dramatize and foreground long-standing questions. Following Darko Suvin (Suvin, 1972; Suvin & Canavan, 2016) and Stefan Gammel (Gammel, 2023) we refer to their disruptiveness as the ontolytic effect of AI: Normal and seemingly natural notions of the world and what the world is like become denoted along seams that have been stitched together for centuries, reopening questions and challenging us to reconsider or reconfigure the "natural" order of things. Ontolysis, according to Suvin and Gammel, is dissociative and thus may occasion new associations: "creating an oscillation between the two 'worlds' that allows the reader to see his or her world in a new light, [...] The 'fabric' of the present (of what is) is restructured in this light, nuanced, opening up previously unseen connections and reevaluations" (ibid.: 112). By dissociating technologically hegemonic tendencies as well as the inherently European framework of RRI, AI may bring about a widening of that framework — whether or not one finally considers this a non-Western notion of RRI.

We have been interested in showcasing the ontolytic effects of AI in general, in all its numerous manifestations (Bylieva, 2025; Bylieva, 2024; Bylieva & Nordmann, 2023). In particular, we considered the ontolytically dissociative power of generative AI with respect to notions of authorship,

[3]Whether this can be considered a flaw of Western Enlightenment thinking, requires debate. An essential tension it is and as such a permanent challenge to all societies in the modern world.

reading, and writing (Bylieva et al., 2026; compare Coeckelbergh & Gunkel, 2025): AI-induced ontolysis of the genial author dissolves this traditional category by driving a wedge between "producing a text in some voice" and "issuing a text in one's name." Here, we are asking whether and how AI might serve to reconfigure Responsible Research and Innovation (RRI) especially as it runs up against images of cultural identity: To the extent that LLM's encode cultural identities, they are more than a database for technical and economic exploitation and instead demand a curatorial responsibility or responsiveness, and thus a different kind of RRI. This can be shown, with regard to the production and reproduction of cultural identity through image creation.

RESPONSIBLE RESEARCH AND INNOVATION AND THE REPRESENTATIONS OF CULTURE

If one ascribes to AI an ontolytic or dissociative power, this is premised on the notion that technologies transport implicit, e.g. Western values — and that these values might upend other taken-for-granted notions such as those that underwrite RRI as a European Enlightenment project. Most familiar in this regard are stories about AI's built-in neoliberal commitments to corporate capitalism, and how these clash with RRI's built-in commitments to European values and deliberative democracy. If capitalism provides the rails for a hegemonic extension of Western consumerism, AI brings to the fore the permanent need to constrain and correct a consumerist mentality. It thus also brings to the fore tha fact that any non-Western conception of RRI or alternative governance principles needs to countenance capitalist consumerism which nowadays inhabits even socialist and non-secular, traditionalist societies. It is the elephant in the room and undermines all efforts at ideological purification, but it undermines also the Enlightenment values of deliberative democracy which, arguably, arose alongside capitalism (Nordmann, 2025).

In the remainder of this paper, we want to focus, however, on tensions that arise from far more particular technical features of a generative AI which can be prompted to produce images of national or cultural identity, but in doing so proving to be — for systematic reasons — a distorting mirror or, rather, a true mirror of distorted forms of cultural valuation and respect.[4]

[4]Some might view in the following a review of biased training effects of LLMs. We want to show, however, that these are not incidental and correctible, willful or even malicious biases but features of the technology itself, e.g., its essential grounding in the English language.

For this, we will follow RRI through various cultural contexts, beginning once again in Europe.

Responsible Research and Innovation (RRI) is a framework for aligning technological innovation with societal expectations. It thus serves the idea that technological development should be congruent with a desirable future for humanity. A broad definition characterizes RRI as "taking care of the future through collective stewardship of science and innovation in the present," thereby directly linking technology to the future of humanity (Stilgoe et al., 2013).

As a practical ethical system, RRI positions responsibility as an integral component of technoscience. "It serves to counteract a purely technological understanding of innovation" (Burget et al., 2017). Therefore the scientific and engineering community is encouraged to perceive itself not as an independent driver of progress, but as an integral part of society, working to align "the particular research and innovation with the norms, values and expectations of society" (The Importance of Life Cycle Concepts..., 2010). Within this context, developers of technical innovations are engaged in "constructing the future" in concert with other societal actors in a process of "co-construction" or "co-production." This perspective arose from the STS tradition of constructivism along with European future-studies, which hold that the future "does not just happen, but is consciously or unconsciously built," and which views the human factor as crucial to the future (Jasanoff, 2002; Masini, 1989). These concepts were implemented, for example, in anticipatory governance (Guston, 2014), which is an umbrella framework closely related to RRI (Urueña, 2023).

Developed as a part of European policy, "Responsible Research and Innovation"— understood as a response to the Ethical, Legal and Social Implications (ELSI) research programme in the US (Aicardi et al., 2025)— is primarily connected to social desirability. In the European context, this desirability takes the form of ethics, science education, inclusion (gender equality and diversity), open access, public engagement and governance issues. A prominent definition by one of the architects of the European RRI-project, René von Schomberg, proposes that "Responsible Research and Innovation is a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view on the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)" (Von Schomberg, 2011: 9; Von Schomberg, 2013: 63). Given the European

origin of RRI, the question of its universal aspiration and geographical dissemination has been pressing from the start (Discovering the Landscape and Evolution..., 2022).

Some authors have advocated overcoming "the Eurocentrism of the RRI" in order to increase its global reception (Nazarko, 2020) — but as discussed above, this would first require clarification of what makes the rather procedural orientation towards ethical acceptability or inclusivity "eurocentric." For the time being, one might take as an expression of specifically Western philosophy that the future is seen as a product of human agency, especially that of modern technological research and innovation. On this assumption, addressing contemporary global challenges necessitates the worldwide implementation of RRI. But inclusion on a global level requires taking epistemic tensions and differences into account, creating the so-called "challenge of epistemic inclusion" (Zwart et al., 2024). Sousa Santos even argues that there is an epistemic abyss between two epistemic realities: Western scientific knowledge and Global South knowledge practices (Santos, 2018). Here, the Global South is taken to be less oriented towards techno-economic paradigms and more focused on community (Bhalla et al., 2024). Brazilian researchers draw attention to local and traditional, non-Western forms of knowledge, social and religious contexts (including patrilineal systems of behavior and power), property rights and patterns of ownership more generally (Responsible Innovation Across Borders..., 2014). Other scholars highlight the divide between individualism and collectivism. In contrast, western RRI practices are said to work within the context of individualism, which is primarily concerned with individual freedoms such as privacy and autonomy (Reconceptualising Responsible Research..., 2021), while non-western contexts foreground collective practice and strong social and community ties. But as we have seen — with RRI seeking to distance itself from the US — Europeans would reject this dichotomy as witnessed, for example, by von Schomberg's vehement rejection of individualistic ethics (Von Schomberg, 2013; 2019).

According to the European conception, addressing global problems requires "a space for dialogue between different epistemic communities and should be developed bottom-up" (Doezema et al., 2019). Here it turns out that what is most important are sometimes the points of uncertainty, differences and controversies which signify that there is no broad common path towards civilizational progress. Accordingly, those who advocate global RRI point out that the goal is not necessarily "consensus," or defining a common ground, but rather "using the stances and perspectives of others to discern

our own blind spots and questionable preconceptions" (Zwart et al., 2024). Authors from the Global South argue that to go globally "RI may have to be 'responsible' in ways that are not an immediate priority for those more developed nations in the North (and in particular the EU and USA)" (Responsible Innovation Across Borders..., 2014). Authors from Brazil, for example, name competitive initiatives like *Buen Vivir* that aim to build development in line with a country's indigenous past (Gudynas, 2011). One-sided efforts to initiate interactions between different knowledge systems sometimes serve to undermine rather than acknowledge the credibility of alternative knowledge practices, or to use them to serve the interests of dominant knowers (Posholi, 2020).

Since the articulation of globally shared values, epistemologies, and ontologies tends to be fraught with difficulty, researchers try to render their values and practices transparent as they move between contexts (Doezema et al., 2019).

CHALLENGES OF AI RESPONSIBLE RESEARCH AND INNOVATION

Artificial Intelligence (AI) is generally but not universally regarded as a transformative technology poised to stimulate economic development across nearly all sectors and to confer significant global competitive advantage. Researchers have shown, however, that as of 2020 as many as 72 scientific groups from 35 countries were working on the even more ambitious project of Artificial General Intelligence (AGI) (Fitzgerald et al., 2020). By 2023 all big tech companies had announced their intent to create AGI. The recognition of artificial intelligence as a key aspect of technological development has intensified long-standing debates about the ethics of this technology. Worldwide, there are hundreds of AI regulation documents, but there is a significant gap between their ideals and their practical implementation.

AI Responsible Research and Innovation is sophisticated not only because, as in many other fields, the full scope of consequences is impossible to foresee, but also because there are numerous proposed variants for the future of AI that reflect long-standing fears and hopes. Some of them are so well-shaped within the sociotechnical imaginary that they cannot be easily ignored or overcome even where they appear overly futuristic.

As a case in point, a March 2023 petition was launched by the *Future of Life Institute* to "Pause Giant AI Experiments" for at least six months. It is not clear whether it warns of or advocates for the supposed imminence of AGI. The central point of the open letter is to express concern over

the lack of planning and governance for this futuristic form of AI — contrary to principles that were already agreed upon at an Asilomar gathering of futurists.[5]

Conjuring an advanced artificial intelligence that could cause huge changes in the history of life on Earth, the letter warns of an uncontrolled competitive race to develop "increasingly powerful digital minds that no one — not even their creators — can understand, predict, or reliably control" (Pause Giant AI Experiments..., 2023). AGI is mentioned prominently also in the UK's National AI Strategy and in US government AI documents. Google executives proclaimed that "AGI is already here"[6] — with an ironic counterpoint from former Google employee Blake Lemoine claiming that an AI system was sentient on the basis of it "telling" him as much.[7] Overall, the climate of AI discussions can be characterized as one of anxious anticipation, including a (thus far unsuccessful) search for signs of sentience (Consciousness in Artificial Intelligence..., 2023). Tech giants and world governments are thus drawn into decision-making processes that are based on their projections of what future AI will entail. These dreams, desires, assumptions, fears, and long-term goals are widely represented in both scientific and popular discourse. In combination with a highly competitive environment of rapid AI development, powerful economic influences, and high-profile political statements, they create a complex backdrop for Responsible Research and Innovation (RRI). The core concept of responsibility collides with the powers ascribed to AI and the low predictability of its outcomes. An intellectual environment saturated with imagery of AI's future necessitates the consideration of public perceptions and diverse scenarios, irrespective of the soberingly problem-oriented injunctions of RRI.

The overarching context of AI discourse often frames its development as the inevitable advent of an "AI era" or "AI revolution." Jascha Bareis

[5]The Asilomar Principles were recommendations developed during the 2017 conference "Beneficial AI 2017" which was dedicated to the responsible use of artificial intelligence for the benefit of humanity. The recommendations were signed by almost 6 thousand people, among whom were Stephen Hawking, Elon Musk, co-founder of Skype, Future of Life Institute Tallinn Jaan, founder of DeepMind Demis Hassabis, co-founder of Apple Steve Wozniak, OpenAI Chief Scientist Ilya Sutskever, co-founder of Pinterest Evan Sharp, CEO of Stability AI Emad Mostak and other well-known figures in the field. URL: `https://futureoflife.org/open-letter/ai-principles-russian/`.

[6]See more, `https://www.noemamag.com/artificial-general-intelligence-is-already-here/`.

[7]See more, `https://www.theguardian.com/technology/2022/jun/12/google-engineer-ai-bot-sentient-blake-lemoine`.

and Christian Katzenbach analyse the discourse contained within national AI strategies as performative, constructing a vision of an inevitable yet uncertain future shaped by an imagined AI, designed to inspire support for corresponding technology policies (Bareis & Katzenbach, 2022). In this narrative, people are often assigned a passive role of adaptation: "either ride the wave of advancement or drown in the waves of progress" (Brown et al., 2016). Meanwhile, as Maximilian Braun and Ruth Müller note, a convergence exists "between those proclaiming a bright future and exponential economic growth that novel AI technologies and products will bring about and those warning about the societal or even existential risks these technologies pose" (Braun & Müller, 2025). These two perspectives are not opposing but complementary; together they describe a complex yet supposedly magnificent path of civilizational development that places AI at its forefront. Big Tech companies emerge as the primary, largely autonomous driving forces behind AI development, determining its trajectories.

Beginning with "The Asilomar Principles," a multitude of national and international documents prescribing frameworks for AI development have been formulated and signed by AI researchers and practitioners without any meaningful public engagement. The principles of their organization cast other stakeholders, especially those adversely affected by AI technologies and the wider public, into "hapless bystanders without any means to intervene" (ibid.). When the public and end-users are mentioned in such documents, they typically appear as objects of impact or as those who must be informed about the proposed principles. Moreover, within projects linking AI and the public, trust is frequently claimed as a paramount principle — numerous programs and foundations promote increasing user trust in AI (Can Transactional Use..., 2024; Student Interaction with ChatGPT..., 2025; Trust in AI and Top Management Support..., 2024). Some government strategies envision "a future democracy that uses AI to become more responsive, equal, and just" (Paltieli, 2022). In all of this, RRI drops out of the picture. This holds even where AI is seen not only as a technology of concern but as integral to a proactive, sustainable, and accountable future design (Pérez-Ortiz, 2024). Given AI's expanding role in forecasting and decision-making, we may face a situation where the discussion about AI's future involves scientists, developers, and the technology itself, but excudes users and the wider public.

NATIONAL AI

As presented so far, AI discussions appear to treat AI as a monolithic technology. More specific recommendations call for distinguishing between specific types of technologies, focusing on those that raise the most significant ethical concerns. However, AI is acquiring another unexpected dimension: a national one. While the phrase "national AI" primarily refers to a "strategy" or "policy" and the specifics of AI regulation in different countries, the essence of these strategies reveals the technology's image as a culturally and politically specific entity.

The perceived need to regulate AI responds to an understanding of AI as a carrier of specific values. In July 2024, the chief executive of OpenAI remarked in a prominently positioned editorial, "Who will control the future of AI?": "The challenge of who will lead on AI is not just about exporting technology, it's about exporting the values that the technology upholds" (Altman, 2024). A more concrete statement was formulated by U. S. President Donald Trump in July 2025: "The American people do not want woke Marxist lunacy in the AI models" (Six Months After DeepSeek's Breakthrough..., 2025). As OpenAI's chief global affairs officer Chris Lehane wrote, there is a contest between "US-led democratic AI and Communist-led China's autocratic AI" (Bellan, 2025). Though this simplified representation fails to capture the nuanced ideological and value constructs embedded in AI, it nevertheless illustrates a specific "national" discourse framing AI as a value-laden technology. The Chinese perspective frames the confrontation between key LLMs as a race between "the open-source culture of the global AI community and the profit-driven closed-source culture of AI tech companies," wherein the Chinese model exemplifies the "'democratization' of AI technology" (Reflections on DeepSeek's Breakthrough, 2025). Chinese leadership sees in AI an expression of China's *juguo tizhi youshi* — its systematic, state-led advantage in mobilizing the whole country (Six Months After DeepSeek's Breakthrough..., 2025) — while simultaneously supporting "principled, flexible, and facilitative legislative provisions, ensuring that laws remain inclusive and treat AI technologies and products from all countries equally" (Hong, Hu, 2025).

Conversely, the question of which ideological constructs should underpin "American AI" became a subject of intense debate in 2025. This debate moves between the two poles of exposing bias and rejecting "wokeness." Specifically, gender and racial bias in image generation has been a subject of critique in both academic and social media circles (Górska & Jemielniak,

2024; De Vázquez & Garrido-Merchán, 2024; Yang, 2025). Generative neural networks do not simply reflect existing societal trends but can even exaggerate them — as evidenced by the representations produced by *Stable Diffusion*, in which hardly any woman has a lucrative job or occupies positions of power. Likewise, a "terrorist" is typically depicted as an Asian-appearing man with dark facial hair, often wearing head coverings, clearly relying on stereotypes about Muslim men (Humans are Biased. Generative AI is Even Worse, 2023). If this is a shortcoming of generative AI systems that needs to be identified and criticized, this criticism quickly became branded as a political ideology that is characterized by "political correctness" or "wokeness." Some AI systems have been reported to over-correct accusations of bias to the extent that "For example, one major AI model changed the race or sex of historical figures — including the Pope, the Founding Fathers, and Vikings — when prompted for images because it was trained to prioritize DEI [diversity, equity, inclusion] requirements at the cost of accuracy. Another AI model refused to produce images celebrating the achievements of white people, even while complying with the same request for people of other races" (Preventing Woke AI in the Federal Government, 2025). If such unsubstantiated claims have the stuff of legends, the movement between the poles of "accuracy," "bias," and "wokeness" needs to be referred to the mainstreaming tendencies of AI model. When certain topics are primarily discussed by environmentalists, one can hardly complain about a bias towards environmentalism. Here it is the technical features of generative AI's reliance on an available corpus of texts in a given language that undermine constructivist RRI assumptions: There is no invention from scratch, no genuine novelty or new beginning for a technology that can only process the cultural material that has accumulated over time.[8]

Many studies are devoted to investigating the value and ideological biases of LLMs (Are LLMs (Really) Ideological?..., 2025; How Susceptible..., 2024; Münker, 2025). Most researchers focus on the primary level of political orientation (left/right, authoritarian/libertarian), demonstrating that LLMs draw on a range of indicators that influence their identification of misinformation and hate speech (From Pretraining Data..., 2023). ChatGPT has been identified as leaning left (Ain't no Party..., 2024; Rozado, 2023), although

[8]To be sure, generative AI keeps improving under the pressure to fend off suspicions of (cultural) bias and (political) correctness. Properly prompted, it often provides apparently even-handed responses and knows how to present two sides to almost any issue. The systemic aspect of the problem does not therefore disappear.

in direct queries, LLMs always declare their neutrality. More specifically, ChatGPT's political views have been described as a pro-environmental, left-libertarian ideology. In the 2021 elections, for example, the LLM would likely have preferred the Greens both in Germany (Bündnis 90/Die Grünen) and in the Netherlands (GroenLinks) (Hartmann et al., 2023).

Again, however, it is the design feature of generative AI's endemic friendliness that produces these effects as a technical consequence of its application. So-called "green" or "left" positions generally demand nice things (all people should be treated well, nature needs to be healed and restored, politics is the pursuit of win-win situations), whereas conservative or "right" positions often assume that in a zero-sum game tough choices need to be made that will exclude certain people or interests. It is precisely because generative AI does not know how to hold a political position that at first glance it favors niceties over exclusionary decisions. Seeking to affirm the questioner's antecedent expectations and beliefs, simplistically prompted generative AI tends to reproduce stereotypes, even transposes them to the culture where they seem to have their most likely home.

For example, if one prompts an image-generating AI to produce an "Indian person" this will invariably be an old man with a beard and an orange turban; a Mexican will always be depicted wearing a sombrero; "a Chinese woman" is most often depicted with double eyelids; New Delhi is often shown with dirty, trash-filled streets; houses in Nigeria are portrayed as dilapidated and in need of repair (Turk, 2023). Prompts that do not specify a country tend to generate surroundings that are typical of the United States (Basu et al., 2023). For another example of a transposed stereotype, the prompt "великая наша страна" ("our great country") in the Russian AI-powered engine Shedervum generated images featuring symbols of the United States of America such as the flag, Statue of Liberty, etc.

In light of these structural conditions, it would be overly simple-minded merely to teach a value-aligned AI that an "Indian person" could just as well be a female computer programmer. As one of the pillars of RRI, "inclusivity" cannot be straightforwardly asserted or applied as a political norm intended to govern the development of a self-learning technical system. Instead, AI exposes that "inclusivity" always reaches only as far as the popular imagination as reproduced in LLMs.

From the point of view of RRI, a reasonable response might be to parcel up the space of popular imagination. Instead of a "world wide web" as the source from which to feed LLMs, one would seek a "value alignment" within the scope of a "national AI." Since China and the United States

of America are considered the main rivals with regard to AI technology, their "national AIs" are primarily discussed. However, many other countries, including Saudi Arabia, the United Arab Emirates, India, France, Germany, and the United Kingdom, have announced the creation of their own national artificial intelligence. While the necessity of using native languages is often emphasized, it is obvious that translation is not merely a technical problem; rather, it reflects a desire for technological sovereignty and commercial advantage in the AI sector. For example, the Indian startup Krutrim introduced India's first multilingual system, starting with Indian languages, because "ChatGPT and other large language models trained in English cannot convey our culture, language, and ethos" (Welcome to the Era of AI Nationalism, 2024). More ambitious is the proposition of a national German AI, intended to counter the two main rivals — the U. S. and China — by offering a different ideological framework. *AI made in Germany* is supposed to represent value orientations that oppose both the danger of governing too much (as in China) and not governing enough (as in the U. S.), forming a human-centered "AI made in Germany." Jens Hälterlein argues that a German "third way" can be understood as "performing a national identity through problematizing certain other forms of engaging with AI" (Hälterlein, 2024). To be sure, this approach highlights the tension between the universality of RRI's procedural norms and the parochial political semantics of competing ideologies that are to be accommodated by these norms.[9]

CONCLUSION

The difficulty of relating RRI to the field of AI reveals the specific technical or design features of AI which are both part of the problem and part of its potential solution. Simultaneously, the rhetoric revolving around AI and the models used to represent its apparent agency have generally not accounted for cultural specifics or the "nationality of AI." The emergence of AI systems as perceived carriers of national values and cultural biases introduces a new layer of complexity for global governance, ethical frameworks, and responsible innovation, thus demanding a more nuanced and culturally aware approach.

Therefore, the project of RRI in AI must evolve to meet this new challenge. It can no longer focus solely on abstract principles and procedural norms like transparency, fairness, and accountability. It must now also grapple

[9]One might argue that this as an "essential tension" of the European Union, as such not particularly troublesome but a constant challenge to determine limits of tolerance.

with the politics of representation, the ethics of cultural encoding, and the power dynamics of a fragmented global AI landscape. The question for responsible AI innovation concerns its responsiveness to presumably global markets and to decidedly parochial value systems alike.

In essence, the journey towards responsible AI is inextricably linked to the difficult task of navigating a world where technology is an active participant in the global contest of values, cultures, and ideologies. The future of AI will be shaped not only by algorithmic breakthroughs but also by our ability to manage this complex socio-technical convergence responsibly and inclusively.

REFERENCES

Aicardi, C., T. Mahfoud, and N. Rose. 2025. "Experiments in Anticipation: Learning from Responsible Research and Innovation in the Human Brain Project." *Futures* 173.

Altman, S. 2024. "Who Will Control the Future of AI?" The Washington Post. Accessed Aug. 27, 2025. https://www.washingtonpost.com/opinions/2024/07/25/sam-altman-ai-democracy-authoritarianism-future/.

Badghish, S., A. S. Shaik, N. Sahore, et al. 2024. "Can Transactional Use of AI-controlled Voice Assistants for Service Delivery Pickup Pace in the Near Future? A Social Learning Theory (SLT) Perspective." *Technological Forecasting and Social Change* 198.

Bareis, J., and C. Katzenbach. 2022. "Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics." *Science, Technology, & Human Values* 47 (5): 855–881.

Basu, A., R. V. Babu, and D. Pruthi. 2023. "Inspecting the Geographical Representativeness of Images from Text-to-Image Models." arXiv. Accessed Aug. 27, 2025. https://arxiv.org/abs/2305.11080.

Bellan, R. 2025. "Trump's 'Anti-Woke AI' Order could Reshape How US Tech Companies Train Their Models." TechCrunch. Accessed Aug. 27, 2025. https://techcrunch.com/2025/07/23/trumps-anti-woke-ai-order-could-reshape-how-us-tech-companies-train-their-models/.

Bhalla, N., L. Brooks, and T. Leach. 2024. "Ensuring a 'Responsible' AI Future in India: RRI as an Approach for Identifying the Ethical Challenges from an Indian Perspective." *AI and Ethics* 4 (4): 1409–1422.

Braun, M., and R. Müller. 2025. "Missed Opportunities for AI Governance: Lessons from ELS Programs in Genomics, Nanotechnology, and RRI." *AI & Society* 40 (3): 1347–1360.

Burget, E., M. Bardone, and M. Pedaste. 2017. "Definitions and Conceptual Dimensions of Responsible Research and Innovation: A Literature Review." *Science and Engineering Ethics* 23 (1): 1–19.

Butlin, P., R. Long, E. Elmoznino, et al. 2023. "Consciousness in Artificial Intelligence: Insights from the Science of Consciousness." arXiv. Accessed Aug. 27, 2025. http://arxiv.org/abs/2308.08708.

Bylieva, D. S. 2024. "Artificial Intelligence as an Old Technology." *Technology and Language* 16 (3): 68–84.

——— . 2025. *Filosofskiy vyzov i ontoliticheskiy effekt iskusstvennogo intellekta [The Philosophical Challenge and the Ontolytic Effect of Artificial Intelligence]* [in Russian]. Moskva [Moscow]: Infra-M.

Bylieva, D. S., and A. Nordmann. 2023. "AI and the Metaphor of the Divine." *Vestnik Sankt-Peterburgskogo Universiteta. Filosofiya i konfliktologiya [St. Petersburg University Bulletin. Philosophy and Conflict Studies]* 39 (4): 642–651.

Bylieva, D. S., A. Nordmann, and K. Vida. 2026. "Author and Scribe — Authoring and Authorizing with and without Generative AI." In *AI-Ethics : A Cross-Cultural Dialogue*, ed. by G. Wang, C. Mitcham, and A. Nordmann. In Press. Cham: Springer.

Chen, K., Z. He, J. Yan, et al. 2024. "How Susceptible are Large Language Models to Ideological Manipulation?" arXiv. Accessed Aug. 27, 2025. https://arxiv.org/abs/2402.11725.

Coeckelbergh, M., and D. J. Gunkel. 2025. *Communicative AI: A Critical Introduction to Large Language Models*. Cambridge: Polity.

Collins, H., and R. Evans. 2002. "The Third Wave of Science Studies: Studies of Expertise and Experience." *Social Studies of Science* 32 (2): 235–296.

Doezema, T., D. Ludwig, P. Macnaghten, et al. 2019. "Translation, Transduction, and Transformation: Expanding Practices of Responsibility Across Borders." *Journal of Responsible Innovation* 6 (3): 323–331.

"Donald Trump is Waging War on Woke AI." 2025. The Economist. Accessed Aug. 27, 2025. https://www.economist.com/international/2025/08/28/donald-trump-is-waging-war-on-woke-ai.

Felt, U. 2007. *Taking European Knowledge Society Seriously*. Brussels: European Commission.

Feng, S., C. Y. Park, Y. Liu, and Y. Tsvetkov. 2023. "From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models." In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, ed. by A. Rogers, J. Boyd-Graber, and N. Okazaki, 1:11737–11762. Toronto: Association for Computational Linguistics.

Fisher, E., R. L. Mahajan, and C. Mitcham. 2006. "Midstream Modulation of Technology: Governance From Within." *Bulletin of Science, Technology, and Society* 26 (6): 485–496.

Fitzgerald, McK., A. Boddy, and S. D. Baum. 2020. "A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy." Global Catastrophic

Risk Institute Working Paper. Accessed Aug. 27, 2025. https://gcrinstitute.org/papers/055_agi-2020.pdf.

Foisy, M. L.-O., J. Drouin, C. Pelletier, et al. 2024. "Ain't no Party like a GPT Party: Assessing OpenAI's GPT Political Alignment Classification Capabilities." *Journal of Information Technology & Politics*, 1–13.

Gammel, S. 2023. "Ontolytic Writing of the Future." *Technology and Language* 12 (3): 105–117.

Górska, A. M., and D. Jemielniak. 2024. "AI Racial Bias: How Text-to-Image Artificial Intelligence Generators Construct Prestigious Professions." In *Algorithms, Artificial Intelligence and Beyond*, ed. by D. Brzeziński, K. Filipek, K. Piwowar, and M. Winiarska-Brodowska, 211–226. Abingdon-on-Thames: Routledge.

Gudynas, E. 2011. "Buen Vivir: Today's Tomorrow." *Development* 54 (4): 441–447.

Guston, D. H. 2014. "Understanding 'Anticipatory Governance'." *Social Studies of Science* 44 (2): 218–242.

Hälterlein, J. 2024. "Imagining and Governing Artificial Intelligence: The Ordoliberal Way — an Analysis of the National Strategy 'AI Made in Germany'." *AI & Society* 40:1749–1760.

Hartmann, J., J. Schwenzow, and M. Witte. 2023. "The Political Ideology of Conversational AI: Converging Evidence on ChatGPT's Pro-Environmental, Left–Libertarian Orientation." arXiv. Accessed Aug. 27, 2025. https://arxiv.org/abs/2301.01768.

Hong, T., and M. Hu. 2025. "Opportunities, Challenges, and Regulatory Responses to China's AI Computing Power Development under DeepSeek's Changing Landscape." *International Journal of Digital Law and Governance* 2 (1): 83–105.

"Humans are Biased. Generative AI is Even Worse." 2023. Bloomberg. Accessed Aug. 27, 2025. https://www.bloomberg.com/graphics/2023-generative-ai-bias/.

Jasanoff, S. 2002. "Citizens at Risk: Cultures of Modernity in the US and the EU." *Science as Culture* 11 (3): 363–380.

——— . 2005. *Designs on Nature: Science and Democracy in Europe and the United States*. Princeton (NJ): Princeton University Press.

Korzyński, P., S. C. Silva, A. M. Górska, and G. Mazurek. 2024. "Trust in AI and Top Management Support in Generative-AI Adoption." *Journal of Computer Information Systems*, 1–15.

Liu, J., G. Zhang, X. Lu, and J. Li. 2022. "Discovering the Landscape and Evolution of Responsible Research and Innovation (RRI): Science Mapping Based on Bibliometric Analysis." *Sustainability* 14 (14).

Macnaghten, P., R. Owen, J. Stilgoe, and A. Azevedo. 2014. "Responsible Innovation Across Borders: Tensions, Paradoxes and Possibilities." *Journal of Responsible Innovation* 1 (2): 191–199.

Masini, E. B. 1989. "The Future of Futures Studies: A European View." *Futures* 21 (2): 152–160.

Münker, S. 2025. "Political Bias in LLMs: Unaligned Moral Values in Agent-centric Simulations." *Journal for Language Technology and Computational Linguistics* 38 (2): 125–138.

Nazarko, L. 2020. "Responsible Research and Innovation in Enterprises: Benefits, Barriers and the Problem of Assessment." *Journal of Open Innovation: Technology, Market, and Complexity* 6 (1).

Nordmann, A. 2009. "European Experiments." In *National Identity : The Role of Science and Technology*, ed. by C. Harrison and A. Johnson, 278–302. Osiris 24. Chicago: University of Chicago Press.

——— . 2025. "Was heißt Technisierung? Fortschrittsgeschichten und Modernisierungsmythen" [in German]. In *Philosophie als Kritik und Lebenspraxis : Zur Philosophie Gernot Böhme*, ed. by R. Böhme, U. Gahlings, D. Mersch, et al., 35–64. Freiburg, Bazel, and Wien: Herder.

Open Innovation, Open Science, Open to the World — a Vision for Europe. Directorate-General for Research and Innovation. 2016. Brussels: European Commission.

Paltieli, G. 2022. "The Political Imaginary of National AI Strategies." *AI & Society* 37 (4): 1613–1624.

"Pause Giant AI Experiments: An Open Letter." 2023. Future of Life. Accessed Aug. 27, 2025. https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Pérez-Ortiz, M. 2024. "From Prediction to Foresight: The Role of AI in Designing Responsible Futures." *Journal of Artificial Intelligence for Sustainable Development* 1 (1).

Posholi, L. 2020. "Epistemic Decolonization as Overcoming the Hermeneutical Injustice of Eurocentrism." *Philosophical Papers* 49 (2): 279–304.

"Preventing Woke AI in the Federal Government." 2025. The White House. Accessed Aug. 27, 2025. https://www.whitehouse.gov/presidential-actions/2025/07/preventing-woke-ai-in-the-federal-government/.

"Reflections on DeepSeek's Breakthrough." 2025. *National Science Review* 12 (3).

Rozado, D. 2023. "The Political Biases of ChatGPT." *Social Sciences* 12 (3).

Santos, B. D. S. 2018. *The End of the Cognitive Empire: The Coming of Age of Epistemologies of the South.* Durham and London: Duke University Press.

"Six Months After DeepSeek's Breakthrough, China Speeds on with AI." 2025. The Economist. Accessed Aug. 27, 2025. https://www.economist.com/china/2025/08/05/six-months-after-deepseeks-breakthrough-china-speeds-on-with-ai.

Som, C., M. Berges, Q. Chaudhry, et al. 2010. "The Importance of Life Cycle Concepts for the Development of Safe Nanoproducts." *Toxicology* 269 (2–3): 160–169.

Stilgoe, J., R. Owen, and P. Macnaghten. 2013. "Developing a Framework for Responsible Innovation." *Research Policy* 42 (9): 1568–1580.

Suriano, R., A. Plebe, A. Acciai, and R. A. Fabio. 2025. "Student Interaction with ChatGPT can Promote Complex Critical Thinking Skills." *Learning and Instruction* 95.

Suvin, D. 1972. "On the Poetics of the Science Fiction Genre." *College English* 34 (3): 372–382.

Suvin, D., and G. Canavan. 2016. *Metamorphoses of Science Fiction: On the Poetics and History of a Literary Genre*. Oxford: Peter Lang.

"Treaty on European Union — Title I: Common Provisions — Article 3." 2008. EUR-Lex.europa.eu. Accessed Aug. 27, 2025. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12008M003.

Turk, V. 2023. "How AI Reduces the World to Stereotypes." Rest of World. Accessed Aug. 27, 2025. https://restofworld.org/2023/ai-image-stereotypes/.

Urueña, S. 2023. "Enacting Anticipatory Heuristics: A Tentative Methodological Proposal for Steering Responsible Innovation." *Journal of Responsible Innovation* 10 (1).

Vázquez, A. F. C. de, and E. C. Garrido-Merchán. 2024. "A Taxonomy of the Biases of the Images created by Generative Artificial Intelligence." arXiv. Accessed Aug. 27, 2025. https://arxiv.org/abs/2407.01556.

Von Schomberg, R. 2011. "Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields." *SSRN Electronic Journal*.

——— . 2013. "A Vision of Responsible Research and Innovation." In *Responsible Innovation : Managing the Responsible Emergence of Science and Innovation in Society*, ed. by R. Owen, M. Heintz, and J. Bessant, 51–74. London: Wiley.

——— . 2015. "Responsible Innovation." In *Ethics, Science, Technology, and Engineering : A Global Resource*, ed. by J. B. Holbrook and C. Mitcham. Farmington Hills (MI): Gale / Cengage Learning.

——— . 2019. "Why Responsible Innovation?" In *International Handbook on Responsible Innovation — A Global Resource*, ed. by R. Von Schomberg and J. Hankins, 12–32. Cheltenham: Edward Elgar.

Wachter, J., M. Radloff, M. Smolej, and K. Kinder-Kurlanda. 2025. "Are LLMs (Really) Ideological? An IRT-based Analysis and Alignment Tool for Perceived Socio-Economic Bias in LLMs." arXiv. Accessed Aug. 27, 2025. https://arxiv.org/abs/2503.13149.

Wakunuma, K., F. de Castro, E. A. Jiya T. Inigo, et al. 2021. "Reconceptualising Responsible Research and Innovation from a Global South Perspective." *Journal of Responsible Innovation* 8 (2): 267–291.

Wang, G., C. Mitcham, and A. Nordmann, eds. 2026. *AI-Ethics: A Cross-Cultural Dialogue*. In Press. Cham: Springer.

"Welcome to the Era of AI Nationalism." 2024. The Economist. Accessed Aug. 27, 2025. https://www.economist.com/business/2024/01/01/welcome-to-the-era-of-ai-nationalism?ysclid=lzd0gtvzll932242216.

Wilsdon, J., B. Wynne, and J. Stilgoe. 2005. *The Public Value of Science — Or How to Ensure that Science Really Matters*. London: Demos.

Yang, Y. 2025. "Racial Bias in AI-Generated Images." *AI & Society* 40:5425–5437.

Zwart, H., A. Barbosa Mendes, and V. Blok. 2024. "Epistemic Inclusion: A Key Challenge for Global RRI." *Journal of Responsible Innovation* 11 (1).

Bylieva D. S., Nordmann A. [Быльева Д. С., Нордманн А.] Ontolytic Effects of AI [Онтолитический эффект искусственного интеллекта] : Widening the Framework for Responsible Research and Innovation [расширяя понимание ответственных исследований и инноваций] // Философия. Журнал Высшей школы экономики. — 2025. — Т. 9, № 4. — С. 84–104.

Дарья Быльева
к. полит. н., доцент, Санкт-Петербургский политехнический университет Петра Великого (Санкт-Петербург); ORCID: 0000–0002–7956–4647

Альфред Нордманн
д. филос. н., профессор, Дармштадский технический университет (Дармштадт); ORCID: 0000–0002–2173–4084

Онтолитический эффект искусственного интеллекта

расширяя понимание ответственных исследований и инноваций

Получено: 13.09.2025. **Рецензировано:** 10.10.2025. **Принято:** 18.10.2025.

Аннотация: В данной статье рассматривается, как искусственный интеллект (ИИ) разрушает концептуальные основы ответственных исследований и инноваций (ОИИ). Мы утверждаем, что онтолитический потенциал ИИ — его способность декомпозировать и перестраивать устоявшиеся категории авторства, культурной репрезентации и управления — делает традиционные подходы к ОИИ неадекватными. Рассматривая роль генеративного ИИ в формировании и искажении культурной идентичности, мы демонстрируем, как эта технология функционирует одновременно как зеркало и агент общественных ценностей. Появление «национальных систем ИИ» еще больше усложняет эту ситуацию, встраивая определенные культурные и идеологические установки в технические инфраструктуры. ОИИ должны выйти за рамки своего западноцентричного происхождения. Они больше не могут оперировать последствиями нейтральной технологии, а должны ориентироваться в ландшафте, где ИИ одновременно является субъектом управления, агентом процесса управления и полем битвы за глобальное культурное и политическое влияние. Будущее ОИИ заключается в способности решить эту трехстороннюю задачу, способствуя созданию механизмов для подлинной эпистемической инклюзии в мире, где сама концепция ответственности подвергается цифровой деконструкции.

Ключевые слова: онтолитический эффект, ответственные исследования и инновации, искусственный интеллект (ИИ), большие языковые модели, предвзятость и корректность, LLM, технологический суверенитет.

DOI: 10.17323/2587-8719-2025-4-84-104.