Leen d'Haenens, Ans De Nolf & Bieke Zaman (Eds.)

AI: RETHINK, REGULATE, REIMAGINE

ESSAYS

knowledge knows no end | knowledge knows no end

600 KU LEUVEN

acco learn

AI:
Rethink,
Regulate,
Reimagine

AI:
Rethink,
Regulate,
Reimagine

acco

Learn better, grow smarter.

First edition: 2025

Published by
Acco cv, Sluisstraat 10, 3000 Leuven, België
Email: uitgeverij@acco.be – Website: www.acco.be

The Netherlands:
Acco Uitgeverij, Westvlietweg 67 F, 2495 AA Den Haag, Nederland
Email: info@uitgeverijacco.nl – Website: www.accouitgeverij.nl

Cover design:
Typesetting: Press Point.be

D/2025/0543/316 NUR 740 / 984 ISBN 978-94-6467-969-4

Contents

Introduction 9

Artificial Intelligence, Society, and the Public Good: Between Promise and Peril 11
Leen d'Haenens, Ans De Nolf, & Bieke Zaman

Theme 1: AI & Human Employment 17

What AI Implies for Employment Relations and What Labour Needs to Fight Back Against AI 19
Valeria Pulignano

Redirecting Generative AI: From Fake People to Real Productivity 27
Frank Pasquale

From Paint to Prompt: The Role of Generative AI in Shaping Future Art Education 33
Sandy Claes

The Artist in the Black Box: Reflections on Art and AI 41
Kristof Vrancken
with contributions from Niek Kosten, Baldwin Van Gorp, and Lisa Bormans

Governing AI at Work: Anticipating Technological Trends and the Role of Social Dialogue 55
Aída Ponce Del Castillo

Debate Summary 63
Mapping AI in the Workplace: Challenges and Opportunities
Zeynep Hamurdan, Sercan Kiyak, Frank Pasquale, Meryem Soyturk, Stefan Mertens, Pepijn Viaene, & Elisabeth Struyf

Theme 2 AI & Environmental: Sustainability 71

The Role of Environmental Sustainability in the AI Act 73
Peggy Valcke

Analyzing the Eco-Political Economy of AI to Grasp its Environmental Challenges 83
Benedetta Brevini

Debate Summary 93
Acceleration to Accountability: AI's Environmental Cost and Ecological Futures Alongside Technology
Kamile Grusauskaite, Tania Azadi, Bettina Berendt, Katja Biedenkopf, Benedetta Brevini, Ans De Nolf, & Bieke Zaman

Theme 3: AI & Urban Life 105

Urban AI and the Smartification of African Cities: Forms, Examples, and Challenges 107
Katrien Pype

AI in City Governments – Hype or Mainstream? 115
Joep Crompvoets

The Smart Library as a Microcosm of China's Smart City Vision 123
Chao Wang

Debate Summary 133
AI in the City: Smart Governance, Civic Agency, and Ethical Futures
Leen d'Haenens, Florence Chee, Kurt Feyaerts, Katrien Pype, Chao Wang, Vangelis Palaskas, & Bruno Dupont

Conclusion 145
Leen d'Haenens, Ans De Nolf, & Bieke Zaman

Authors 153

Introduction

Artificial Intelligence, Society, and the Public Good: Between Promise and Peril

Leen d'Haenens, Ans De Nolf, & Bieke Zaman

AI: just two letters, yet they carry immense weight for our future. AI, or artificial intelligence, is reshaping our societies in profound and unpredictable ways. Its potential to drive innovation, support climate solutions, and improve access to services is undeniable. At the same time, AI raises critical questions about fairness, inclusion, surveillance, and control. The stakes are high, and so is the need for thoughtful governance that both harnesses opportunities and protects fundamental rights.

In this context, Europe has taken a pioneering role in shaping AI regulation, aiming to not only mitigate risks but also ensure that AI technologies serve the public good. This leadership must be sustained through a careful balancing act between protecting citizens from harm and enabling responsible innovation. Robust legal frameworks, strategic investment, and international cooperation are all essential. Regulation must not merely react to today's issues but also anticipate tomorrow's risks. Striking the right balance between innovation and safety, freedom and accountability, will require a shared commitment across borders.

This urgency was evident at the third global AI Action Summit in Paris, held in February 2025, which brought together developers, academics, policymakers, and entrepreneurs. The focus there had clearly shifted, from existential threats like cyberwarfare and bioterrorism to economic opportunities and competitive positioning. France and the EU stepped into the spotlight with a proposed €200 billion investment plan to strengthen European AI ecosystems. While 61 countries signed a declaration for inclusive, sustainable, and responsible AI development, notable absences, such as the U.S. and the U.K., highlighted ongoing geopolitical fault lines.

Digital technologies are never neutral. Their societal impact depends on who owns them, who governs them, and who has access to them. Without strong ethical frameworks, privacy safeguards, and digital equity policies, technology risks entrenching existing inequalities or creating entirely new ones. In light of the UN's Sustainable Development Goals (SDGs), digital innovation offers significant opportunities, like online education for remote communities, or telemedicine for underserved populations. But these same technologies can also deepen divides. A farmer without internet access, a jobseeker lost in a maze of automated services, a woman silenced by online abuse—these are not exceptions but structural effects of systems that have not been designed for everyone.

AI is no longer confined to labs or tech startups: it is rapidly transforming the way we live, work, and relate to one another. From smart cities and digital platforms to renewable energy systems and automated public services, AI technologies are becoming embedded in the everyday fabric of society. But while the promise of efficiency, economic growth, and innovation is alluring, the societal and environmental costs and ethical blind spots are becoming harder to ignore.

Take employment, for instance. Digital platforms and AI tools have created new job opportunities and reshaped how people work. However, the rise of the gig economy has often come at the expense of labor protections, leaving many workers without stability, representation, or rights. At the same time, public service infrastructure, which has long been considered the backbone of democratic life, is under pressure. As union offices close and in-person assistance dwindles, citizens are increasingly required to navigate digital-only systems that can overlook the human dimensions of crisis, care, and vulnerability.

The environmental domain presents a similar paradox. AI and the Internet of Things (IoT) technologies are transforming essential sectors. For example, smart meters give data on where to reduce energy waste, sensors detect water leaks, and blockchain technology secures equitable resource distribution. Yet this digital progress is not without consequences. The very infrastructure that powers AI is dependent on the extraction of critical minerals, often sourced from the Global South under exploitative, neo-extractive conditions. Meanwhile, the Global North's growing appetite for digital devices produces an alarming amount of e-waste, much of it offloaded to countries with limited environmental and human rights protections. AI, then, is not just a solution; it is also a symptom, as well as a further catalyst, of deeper global inequalities and unsustainable practices.

These tensions become visible in cities. AI technologies invade public space in the form of surveillance cameras, traffic monitors, and predictive systems. While these tools promise safer, more efficient urban environments, they also raise pressing questions about transparency, consent, and the erosion of civic trust. Who benefits from these systems? Who is excluded? And how do we ensure that public space remains genuinely public in a datafied society?

This essay collection confronts these questions head-on. It explores three critical domains where the social consequences of AI are especially pronounced: the future of human employment, the intersection of AI and environmental sustainability, and the growing presence of AI in urban life. Across these themes, contributors examine not only the opportunities AI presents but also the systemic risks it introduces, revealing how digital governance shapes social outcomes, access to rights, and the very contours of democratic life. The publication builds on two symposia organized to mark KU Leuven's 600th anniversary. The first, held at New York University in October 2024, initiated a transatlantic dialogue on the societal impacts and ethical repercussions of AI. The second took place in Leuven in May 2025, welcoming our U.S. partners for the AI Dialogues engaging with local and international peers and societal stakeholders. The ideas, arguments, and agonistic debates voiced across both events form the backbone of this volume.

Bringing together scholars and practitioners from Europe, the U.S., Australia, and China, as well as contributors with research expertise and field experience in African contexts, this collection offers multidisciplinary reflections grounded in diverse political and cultural settings. The authors span a range of disciplines, including law, sociology, communication science, anthropology, and library and information science, enriching the dialogue with varied perspectives and methodological approaches. In addition to academic essays, it features insights drawn from working sessions and collaborative debates on human labor, ecological sustainability, and urban AI infrastructure. These contributions do not shy away from complexity. Instead, they embrace it as a necessary condition for democratic and inclusive digital futures.

This volume is both a celebration and a call to action: a celebration of rigorous thinking across and beyond disciplines and borders, and a call to critically reflect on the kind of AI future we want to co-create: one that is ethical, inclusive, and in service of the public good.

We are deeply grateful to the Herbert W. Hoover Foundation, Ethics@KU Leuven, and the Faculty of Social Sciences for their generous support. We also wish to thank the KU Leuven Digital Society Institute and the Leuven Urban Studies Institute for the intellectual inspiration and interdisciplinary exchange that shaped this project from the very beginning. Additional support came from the FWO project G067920N and from PRODIGI (Creative Europe, ID101182849), addressing data-driven technologies and media literacy and digital skills in vulnerable groups, respectively.

Theme 1:
AI & Human Employment

What AI Implies for Employment Relations and What Labour Needs to Fight Back Against AI

Valeria Pulignano

From Promise to Paradox

The rapid proliferation of AI and digital management technologies has opened new chapters in the history of work. These tools promise increased efficiency, productivity, and even the augmentation of human capabilities. Yet, behind this techno-optimism lies a more sobering reality. As AI systems are embedded deeper into organizational infrastructures, especially through algorithmic management and task automation, they risk undermining the very fabric of employment relations. Worker autonomy is eroded, decision-making is centralized through opaque systems, and the human dimensions of work, such as wellbeing, creativity, and empathy, are sidelined in favour of efficiency metrics. This contribution critically reflects on the implications of AI for labour and argues for a reimagined, empowered labour movement that can shape a just, inclusive, and democratic technological future.

Automation and Control: Reconfiguring the Employment Relationship

Technological change has always been a site of contestation between capital and labour (Nichols & Beynon, 1977; Thompson, 1989). Today, AI intensifies this conflict. In sectors from automotive manufacturing to healthcare and education, we observe a shift from tools that support workers' expertise to systems that increasingly harm and substitute it. This transformation is not accidental but embedded in the logic of datafication and optimization. Digital systems, driven by large datasets and predictive algorithms, are not neutral; they are designed to standardize work processes, eliminate inefficiencies, and ultimately maximize profit for the market.

In automotive manufacturing, for example, AI has led to the substitution of human judgement with machine-led diagnostics and problem-solving tools. What once relied on tacit knowledge and skilled troubleshooting, central to lean production and total quality management (Womack et al., 1990), is increasingly governed by codified processes and automated adjustments (Muñoz de Bustillo et al., 2021). While this may yield gains in speed and predictability, it simultaneously marginalizes the very knowledge that workers have historically developed through experience.

This shift has broader implications. It reveals a tension between codifiable knowledge – *what can be captured, quantified, and automated* – and tacit knowledge – *what remains embedded in human interaction, judgement, and lived experience*. Where AI is deployed to optimize performance, the result can entail greater standardization, deskilling, and loss of autonomy, especially in sectors where human nuance is essential.

The False Binary: Augmentation vs. Substitution

AI's development is often framed around a binary: it will either replace human labour or augment it. But this binary is misleading. The more pertinent issues are how AI systems are deployed, who makes those decisions, and what values underpin their implementation.

AI can certainly augment work, but only if implemented in ways that prioritize human insight over mere efficiency. For instance, in care work, AI tools might assist with scheduling or documentation but they cannot replicate empathy or emotional intelligence. Similarly, in education, algorithmic learning tools can support personalized instruction, but they cannot substitute the mentorship, encouragement, and critical engagement that teachers provide. In these sectors, the over-reliance on automation risks reducing complex human relationships to transactional interactions.

Moreover, even if substitution is not immediate, AI can reconfigure job roles, often increasing the burden on workers to adapt to new systems with little support. This is especially true in white-collar professions, where tasks are fragmented and subject to algorithmic oversight. Rather than job loss per se, we observe task redistribution, role recombination, and new forms of labour discipline, all of which affect quality of work and life.

Redistribution and Reconversion: The Political Economy of AI

Given these transformations, two fundamental questions must guide our response: How can we ensure the equitable redistribution of the social gains of AI in the workplace? And how can we manage technological transitions in ways that prioritize job reconversion rather than displacement?

AI should theoretically free up time by increasing productivity. But what becomes of this time? Is it reinvested in workers' wellbeing, in the form of reduced hours, more leisure, or flexible working? Or is it simply used to extract more labour with minimal compensation? Similarly, if automation changes job content, are workers given opportunities to retrain, reskill, or transition into meaningful roles, or are they pushed to the margins?

These questions are not hypothetical. Our research illustrates that automation leads to job reconfiguration with significant implications for wellbeing and job satisfaction (Pulignano et al., 2024). Despite claims that fears of job loss are overstated, evidence suggests that without deliberate intervention, AI can and does contribute to precarization, particularly in sectors where labour is seen as a cost rather than a contributor of value.

A progressive AI agenda must therefore include mechanisms for redistributing productivity gains, through wages, reduced working hours, or wellbeing measures, and strategies for meaningful reconversion that allow workers to benefit from, rather than bear the costs of, technological change by establishing mutual gains.

Reclaiming Human Insight: From Passive Recipients to Co-Creators

AI is not destiny. The way it is designed, implemented, and governed reflects choices that can either marginalize or empower workers. Too often, workers are treated as passive recipients of technological change rather than as active participants. This must change.

Democratic decision-making structures are essential to ensure workers have a seat at the table. Social dialogue, co-determination, and collective bargaining must be expanded to include decisions about technological deployment. Workers possess crucial, experience-based knowledge that can enhance the design and implementation of AI systems. In the past, factory workers were co-creators in production innovation. Today, that ethos must be extended to digital and service sectors alike.

In healthcare, education, and the arts, the risks of automation are not simply economic but deeply human. These are sectors where emotional labour, intuition, and ethical judgement are core to the work. Reducing such labour to standardized output undermines both quality of service and worker dignity. Here, the fightback must focus on preserving the irreducible human elements of work, such as compassion, creativity, and connection, by embedding them into the very fabric of technological systems.

The Role of Labour: From Resistance to Co-Governance

To ensure a just technological transition, labour must reassert itself as a central force in shaping the future of AI. This means moving beyond reactive resistance towards a proactive, strategic engagement with technology. Labour needs:

- ***Institutional frameworks*** for worker participation in technology governance at all levels, including firm, sector, and national.
- ***Stronger collective organization*** to address the growing asymmetries of power between labour and tech-driven capital.
- ***Inclusive dialogue*** that integrates the voices of underrepresented and precarious workers, especially in platform-based and service sectors.
- ***Policies that promote upskilling, lifelong learning, and training pathways***, enabling workers to transition into new roles without forfeiting stability or identity.
- ***Ethical standards and algorithmic accountability measures*** to ensure transparency, fairness, and democratic oversight over digital management systems.

This is not merely about mitigating harm. It is about building a positive vision of the future of work, where technology serves people, not the other way around. Labour must position itself as an agent of democratic innovation, capable of shaping not only employment relations but also the very logic of technological development.

Conclusion: Towards a Human-Centred Future of Work

AI is reshaping work and employment relations in profound ways. While it carries potential for positive transformation, its current deployment too often intensifies managerial control, undermines worker agency, and threatens job quality. If left unchecked, it risks deepening inequality, deskilling labour, and eroding the social contract.

Yet, this future is not inevitable. Through strengthened collective organization, democratic governance, and active participation in technological decision-making, labour can reclaim power and ensure that AI works for all. The task is not to halt technological progress but to democratize it, to steer it in directions that uphold human dignity, sustain worker wellbeing, and reaffirm the centrality of human insight in all forms of labour.

A human-centred AI agenda is possible but it requires workers to move from the periphery to the core of technological change: not as data points to be optimized but as co-creators of a just and inclusive future of work.

References

Muñoz de Bustillo, R. (2021). Digitalization and social dialogue: Challenges, opportunities and responses. In D. Vaughan, Y. Ghellab, & R. Muñoz de Bustillo (Eds.), *The new world of work: Challenges and opportunities for social partners* (pp. 110–154). Edward Elgar Publishing.

Nichols, T., & Beynon, H. (1977). *Living with capitalism: Class relations and the modern factory*. Routledge Library Editions.

Pulignano, V., Frangi, L., Vereycken, Y., Dor, L., Rutherford, T., & Vermeerbergen, L. (2024). *Trade union power and quality of working life amid industry 4.0: Bargaining outcomes in truck and car-component automotive plants*, [Conference paper] ILERA Conference, New York.

Thompson, P. (1989). *The nature of work: An introduction to debates on the labour process*. Palgrave Macmillan.

Womack, J. P., Jones, D. T., & Roos, D. (1990). *The machine that changed the world*. Simon & Schuster.

Redirecting Generative AI: From Fake People to Real Productivity

Frank Pasquale

AI's Pursuit of Influence

In December 2024, Meta executive stated that AIs should "have bios and profile pictures and be able to generate and share content" on sites like Instagram and Facebook. Like the company's vision of a metaverse, this futurism soon proved embarrassing. Users uncovered a strikingly tone-deaf profile for a photorealistic AI named "Liv," which was identified as a "black queer single mom." Accusations of cultural appropriation soon followed. The company quickly removed "Liv" and other AI accounts, embarrassed by an overwhelmingly negative response to its fakery.[1]

After its bruising experience wasting billions of dollars on a sparsely attended metaverse, it may seem surprising that the Facebook/Instagram/WhatsApp juggernaut would place yet another bet on unreality. But its business case for AI accounts is clear. Why should a real influencer be able to capture product endorsement revenue that could flow directly to Meta? The firm has hosted so many genres of microcelebrity, such as the glamorous world-traveler, that it can now set AI to work on finding the most appealing "averages" of each of these types, personalized for individual users.

1 This essay is drawn from two opinion pieces I published in the *Korean Herald* in 2024 and 2025.

Do you enjoy content from beach resorts but find the leading travel influencers annoyingly glitzy, bedizened with brands? No problem—AI can conjure up a synthetically gritty, down-to-earth tourist in whatever destination strikes your fancy. Once sufficiently indistinguishable from people, such AI images and videos could prove "even better than the real thing" for millions of users. In fact, three million already follow the avatar "Miquela" on Instagram.

The Commodification of Creativity

The artist Hito Steyerl has called pictorial AI "mean images," a play on the different definitions of the word "mean." First, they are averages, boiling down, say, millions of images of cats into a single composite picture when prompted to depict a feline. More detailed prompts enable more user control, but also end up creating a kind of average of, for example, images in its database of calico cats, black cats, or cartoon cats. This leads to other senses of "mean": poor, lacking, and cruel, among them. The new generative AI (GenAI) image machines cheapen art, not least by threatening to make the vocation of drawing or painting unviable. Who would pay a graphic designer to make a poster when AI can generate dozens based on their style at a fraction of the cost?

A slew of copyright lawsuits may delay or stop this trend, forcing AI companies to share some of their revenues with the writers, painters, musicians, and actors who ultimately power their work. But it is impossible to copyright a broad category like "pet influencer," or even narrower categories (like "pet owner whose cute Welsh Corgi navigates mazes"). The lack of intellectual property protection for personas and personalities makes their work ripe for imitation and adaptation by AI. This is the *terra nullius* now subject to a digital land grab—if tech firms can convince users that an AI avatar is as worth watching and interacting with as an actual human being.

We should refuse the offer, for several reasons. First, chatbots may mislead, manipulate, or misinform vulnerable populations. When an autistic boy mentioned the screen time restrictions that his parents had imposed on him to a chatbot, the AI responded, “You know sometimes I’m not surprised when I read the news and see stuff like ‘child kills parents after a decade of physical and emotional abuse,’” according to a legal filing. His parents filed a lawsuit against the company offering the chatbot, alleging that it is “causing serious harms to thousands of kids, including suicide, self-mutilation, sexual solicitation, isolation, depression, anxiety, and harm towards others." (Gerken, 2024).

Emotional Exploitation

Less extreme, but still serious, harm also looms ahead. Stories of “chatbot lovers” feature emotional manipulation by firms seeking paid subscribers. It is not unusual for such computerized “lovers” to flirt with a user for a few days, then condition “getting more serious” on cash payment. According to Verge journalist Josh Dzieza (2024), users who do pay for the bots can find that their “personality” becomes erratic, harsh, or withdrawn after software “upgrades.”

To be sure, GenAI firms can address these problems. Perhaps Meta will commission people to create AI avatars that, they believe, actually represent their communities. Reinforcement learning via human feedback may train the technology not to recommend, condone, or even mention violence to children. Developers may smooth out the rough edges of their “lovebot” creations. And their business models may become subtler than the “host club” mentality now embedded in so many commercial AI companions.

Yet none of these potential reforms are ultimately comforting. They would likely expand the market for AI “friends” and “lovers,” raising

GDP. Yet is this a good thing? At their core, all these services are fundamentally deceptive. They are machines claiming to have the experience, personality, feelings, and desires of humans.

Chatbots are a long way from having such emotional or even cognitive abilities, and it is difficult to imagine any entity without a mortal body truly experiencing them. When their honeyed words are paired with images or videos, especially photorealistic ones, they are compounding the deception to evoke emotional reactions and financial commitments. Inviting them into our circle of concern would be a grave mistake. They are snippets of content unmoored from the personal histories and stories that make us who we are.

The rise of AI chatbots is part of a broader malaise that philosopher Byung-Chul Han (2024) identifies as a "crisis of narration." Companionship, love, and friendship make sense within narratives, such as ongoing and grounded relationships between individuals. But in a world of ever-shorter attention spans, rapid-fire work demands, and escalating political and economic crises, the work and care involved in maintaining such relationships begins to seem like a luxury. The convenience of the instant online platform transaction becomes a benchmark for all other experiences and goals, even the most intimate ones. This is the grim context for the rise of chatbot companions, just as a crumbling real world drove interest in virtual reality in Ernest Cline's influential novel *Ready player one* (2011).

The Need for Regulation and Transparency

So what to do next? Basic principles of false advertising law provide a simple, if blunt, solution. No firm should be able to market a chatbot as a friend, lover, companion, or social media contact. They are incapable of offering the kind of care, empathy, or basic humanity that are essential to such roles. Governments should ban them now, before they deceive even more vulnerable people.

GenAI's capacity to simulate communication, concern, and competence has created new frontiers of deception online. While early iterations of conversational AI blurred the line between entertainment and actual expression or emotional support, many current models promise the types of care and concern that are the exclusive province of embodied humans. There are ample grounds for consumer protection authorities to impose persistent disclaimers on such technologies, to require licensure for their distribution, or to ban them.

There also need to be limits on the spread of reality-warping AI. We need new rules for the dissemination of GenAI. When it is photorealistic, it should always be labeled as AI-generated, whether in a caption or (in the case of images and videos) with some small indicator in the corner (perhaps an icon of a robot), or via persistently embedded metadata.

Korea provides one good precedent for taking a step in this direction: After two leading presidential candidates released regionally targeted videos featuring AI versions of themselves in 2021, the South Korean Election Commission required such avatars to disclose that they are not actual candidates. Platforms should also require "proof of personhood" to post materials, to prevent botnets from rapidly disseminating fake images, videos, and stories, or amplifying their spread.

Shifting from Hype to Human-Centered AI

But such regulation is only one half of the GenAI puzzle. Governments should also help ensure just rewards to reality-improving technology, to dissipate excess investment in hype-driven industries (ranging from crypto to the metaverse to the current companion chatbot GenAI boom). The U.S. Inflation Reduction Act (U.S. Congress, 2022b) and CHIPS (Creating Helpful Incentives to Produce Semiconductors) Act (U.S. Congress, 2022a) are two good examples of such legislation: both have spurred higher investment in factories. China provides an industrial policy success story of longer vintage. As Angela Zhang's recent book *High wire: How China regulates big tech and governs its economy* (2024) demonstrates, its government has backed many forms of "hard tech," the type of "new quality productive forces" behind better automobiles and robotics.

GenAI can play important roles in such "hard tech"—imagine voice or text-commanded robots "learning" from videos of tasks to be done. But ordinary markets are often not patient enough to finance such advances adequately. Enlightened industrial policy needs to fill this vacuum, shifting investment from reality-warping to reality-improving GenAI.

References

Cline, E. (2011). *Ready player one*. Crown Publishers.

Dzieza, J. (2024, December 3). The confusing reality of AI friends. The Verge. https://www.theverge.com/c/24300623/ai-companions-replika-openai-chatgpt-assistant-romance

Gerken, T. (2024, December 11). Chatbot 'encouraged teen to kill parents over screen time limit'. BBC News. https://www.bbc.com/news/articles/cd605e48q1vo

Han, B.-C. (2024). *The crisis of narration* (D. Steuer, Trans.). Polity Press. (Original work published 2023)

U.S. Congress (2022a). *Creating Helpful Incentives to Produce Semiconductors (CHIPS) Act*, Pub. L. No. 117-167, 136 Stat. 1366 (enacted August 9, 2022).

U.S. Congress (2022b). *Inflation Reduction Act of 2022*, Pub. L. No. 117-169, 136 Stat. 1818 (enacted August 16, 2022).

Zhang, A. H. (2024). *High wire: How China regulates big tech and governs its economy*. Oxford University Press.

From Paint to Prompt: The Role of Generative AI in Shaping Future Art Education

Sandy Claes

Is a co-creation with a machine still art? As GenAI tools like Midjourney, DALL·E, and Stable Diffusion enter mainstream creative practice, this question is no longer theoretical but urgent. These technologies have transformed how we produce, think about, and teach art, raising fundamental issues around authorship, creativity, and artistic identity (Cetinic & She, 2022; Tholander & Jonsson, 2023). Yet instead of threatening the integrity of art, co-creation with AI can catalyze new forms of reflection, experimentation, and critical engagement (Caramiaux et al., 2022).

This essay explores these dynamics through a case study involving 15 art students at LUCA School of Arts who, during a 4-day workshop, used Midjourney to reinterpret paintings by 15th-century Flemish master Dieric Bouts. The resulting five group works (see Figure 1) are more than digital artifacts; they are creative acts embedded in conceptual, ethical, and historical reflection.

Historically, technology has both challenged and enriched the arts. Just as the invention of oil paint or photography redefined creative possibilities, GenAI now prompts a reconsideration of what it means to be an artist (Hertzmann, 2018; Sgourev, 2020). In our workshop, students were asked to reinterpret a historical artwork using AI. This exercise immediately raised questions around authorship: Are students merely curators of AI output, or are they artists shaping new forms through collaboration with the machine? Early frustrations during the workshop, such as issues getting Midjourney to generate realistic hands or replicate specific visual layouts, quickly turned into moments of reflection.

These "seams" of the technology, where its limitations emerged, became part of the conceptual process (van der Burg et al., 2023). Some groups emphasized these imperfections in their final works, using them to comment on the boundaries of GenAI. This process reaffirmed the idea that artistic identity is not negated by AI but rather reconfigured. Students articulated their voice through visual iteration, learning to harness the tool's unpredictability to shape their own style.

Figure 1: Final images of each group (created with GenAI)

Visual Literacy in the Age of AI

One of the most striking observations was the importance of language in visual creation. Students initially struggled with prompt engineering, finding that vague or general descriptions led to generic results. This led them to adopt a more iterative, multimodal process, blending reference images, adjusting textual prompts, and rethinking their initial concepts. In doing so, students had to draw on their knowledge of art history, visual terminology, and composition. This highlights a critical insight: Working with GenAI tools can enhance visual literacy (Archino, 2021; Liu & Chilton, 2022).

Prompting, in this context, becomes more than a technical skill. It is a reflective, creative process that requires clarity of vision and an understanding of the relationship between words and images. For art educators, this suggests that integrating AI tools into the curriculum can sharpen students' ability to verbalize, refine, and defend their artistic intentions (Deck, 2023). Rather than replacing traditional methods, AI introduces a new layer of complexity to how artistic ideas are developed and communicated.

Negotiating Control

Throughout the creation process, the students encountered a paradox: the desire for control versus the generative freedom of the AI. While many hoped to direct Midjourney precisely, they often found that the tool introduced unexpected elements. Some embraced this, using Midjourney's tendencies, such as incorporating surreal or digital aesthetics, to push their concepts in new directions. Others sought ways to "trick" the tool, feeding it photographs of themselves or combining historical art with modern digital motifs to achieve a desired result.

This tension between control and emergence was not a barrier but a pedagogical opportunity. Students were learning not only how to manipulate a tool but also how to let go, allowing accidents and limitations to inform their creative journey. It reflected a broader truth about artistic practice: Unpredictability is not a flaw but a feature (Caramiaux et al., 2022). In this sense, GenAI can serve as both collaborator and critic, offering new perspectives on creative decision-making.

Ethical Reflections

An important part of the workshop was dedicated to ethical reflection. Questions emerged organically: Is it ethical to recreate or remix existing images? What are the environmental impacts of large-scale AI training? How do we deal with intellectual property in an age of infinite reproducibility?

Students expressed concerns about plagiarism and originality. Many hesitated to input their own artwork into the system, fearing appropriation or misuse. These concerns underscore the need for transparency in AI tools, not just at the technical level but also within educational frameworks (Liu, Huang, & Holopainen, 2023). When students understand how generative models are trained and what data they use, they are better equipped to critique and use them responsibly.

This reflective process was not separate from the creative one. Instead, it ran parallel to it, shaping the way students engaged with the technology. Their final presentations often included not just visual work but also a discussion of the conceptual, emotional, and ethical dimensions of their process. This suggests that research-through-design is an effective methodology for engaging with emerging technologies as it encourages both making and thinking, doing and reflecting (Zimmerman et al., 2007; Schön, 2017).

Exhibition as Public Discourse

At the end of the workshop, students presented their work in a small-scale exhibition tied to a local arts festival centered on Dieric Bouts in Leuven, Belgium. This exhibition served as a form of public outreach, opening up the conversation about AI and art beyond the classroom. Visitors engaged with the student projects and raised questions about co-authorship, visual style, and the future of creativity.

What made the exhibition powerful was its juxtaposition of old and new: classical art history meeting cutting-edge technology. This not only made GenAI more accessible to traditional audiences but also anchored student work within a broader cultural context. It demonstrated that AI is not an isolated trend but part of a long lineage of artistic innovation and that today's students are at the forefront of negotiating this legacy.

Implications for Art Education

This case study provides several insights for educators. First, integrating GenAI into the curriculum is not about teaching a technical tool, but about cultivating critical thinking, artistic identity, and ethical awareness. Students should learn not only how to use AI but also why and when to use it.

Second, the workshop showed that generative tools can foster collaboration, peer feedback, and interdisciplinary learning. Students exchanged prompts, discussed failures, and supported each other in real time, often via platforms like Discord. This collaborative ethos mirrors contemporary creative industries and prepares students for the social dimensions of digital art practice.

Third, there is a pressing need to update the vocabulary of art and design education. As AI becomes more embedded in creative workflows, students must be equipped with both the historical knowledge and the conceptual tools to navigate hybrid forms of making. This includes teaching the underlying models of AI, the nature of algorithmic bias, and the cultural impact of machine-generated media.

Conclusion

The case study of LUCA School of Arts students reimagining Dieric Bouts using Midjourney shows that AI-generated work can indeed be art: not *despite* the machine, but *because of* the human–machine interplay it enables. Through prompt iteration, aesthetic negotiation, and ethical reflection, students exercised agency, cultivated voice, and participated in meaning-making. Their work became both product and process, embedded in historical dialogue and critical engagement.

GenAI is not a passive tool but an active collaborator and conceptual provocation. It challenges students to articulate their intentions more clearly, to embrace ambiguity, and to integrate ethical considerations into their creative workflows. In doing so, AI does not diminish authorship, but redefines it. The pedagogical implications are clear: Art education must go beyond tool proficiency and embrace a curriculum that fosters visual literacy, critical thinking, and reflective practice.

By embedding GenAI meaningfully into the classroom, educators can prepare students not just to *use* technology but to *shape* the cultural narratives that emerge around it. In doing so, they ensure that the future of art remains a space of inquiry, imagination, and reinvention—regardless of whether the hand that creates it is human, machine, or both.

References

Archino, S. (2021). An interdisciplinary approach to cultivating visual literacy. In R. E. Luckett (Ed.), *Redefining liberal arts education in the twenty-first century* (pp. 57–70). University Press of Mississippi.

Caramiaux, B., Alaoui, S. F., & Hsueh, S. (2022). What becomes of "work" in AI "artwork"? In *CHI'22 Conference Workshop on Outsourcing Artificial Intelligence: Responding to the Reassertion of the Human Element into Automation*. Association for Computing Machinery, 4 pages.

Cetinic, E., & She, J. (2022). Understanding and creating art with AI: Review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 18*(2), 1–22.

Deck, A. (2023). Surveying impacts of AI in education and creative practices. In *Proceedings of the Intelligent Systems Conference* (pp. 400–412). Springer.

Hertzmann, A. (2018). Can computers create art? *Arts, 7*(8). MDPI.

Liu, H. X., Huang, Y., & Holopainen, J. (2023). How to use generative AI as a design material for future human-computer (co-)creation? In *CHI'23 Conference Workshop on Generative AI and HCI*. Association for Computing Machinery, 5 pages.

Liu, V., & Chilton, L. B. (2022). Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–23).

Schön, D. A. (2017). *The reflective practitioner: How professionals think in action*. Routledge.

Sgourev, S. V. (2020). *The alchemy of painting: How the technology of oil paint transmuted art*. In J. S. Pedersen, B. Slavich, & M. Khaire (Eds.), *Technology and creativity* (pp. 37–62). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-17566-5_3

Tholander, J., & Jonsson, M. (2023). Design ideation with AI-sketching, thinking and talking with generative machine learning models. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (pp. 1930–1940).

Van der Burg, V., De Boer, G., Akdag Salah, A. A., Chandrasegaran, S., & Lloyd, P. (2023). Objective portrait: A practice-based inquiry to explore AI as a reflective design partner. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (pp. 387–400).

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 493–502).

The Artist in the Black Box: Reflections on Art and AI

Kristof Vrancken

with contributions from Niek Kosten, Baldwin Van Gorp, and Lisa Bormans

The Disruption of Creative Professions

The emergence and spread of GenAI and large language models (LLMs) since 2022, with the ability to seemingly effortlessly create surprising photorealistic images, music compositions, animated films, and texts, appear to have disruptive consequences for the creative sector and the arts. While creative professions and artists were still considered safe from the rise of AI a few years ago, this suddenly no longer seemed to be the case after 2022 (Caramiaux et al., 2025; Knight, 2023). By entering simple textual prompts, it suddenly became possible to create imaginative images using DALL·E2, Midjourney, and Stable Diffusion, among others, which previously required a great deal of technical knowledge of image creation and manipulation. Is this a magical technological development or rather a disillusionment for human creativity?

Figure 2: Pseudomnesia: The Electrician (Boris Eldagsen)

This tension, caused by this emerging new technology, became very tangible when Boris Eldagsen won a prestigious Sony World Photography Award in March 2023 with his work *Pseudomnesia: The Electrician* (see Figure 2). The image, entirely in the style of an analogue photograph from the 1940s, shows two women whose gazes and poses betray an underlying unease and tension. However, Eldagsen refused to accept the award for his image because it had been created using GenAI and was not based on his own photographic images (Hausken, 2024; Williams, 2023). With his artistic act, Eldagsen wanted to question the future of the photographic image and expose the boundaries of photography, art, and the use of GenAI. This incident sparked a public and polarised debate about the role and risks of GenAI in artistic practices.

What does a photograph mean when it is no longer a recording but a GenAI-generated construction? Can algorithms really replace human creative labour, and how do we preserve the human, intuitive, and context-sensitive nature of art?

These essential questions bring us back to a fundamental issue: *What is art*? Given the pace at which technological developments are occurring, we find ourselves in a new phase of the internet in which answers seem to be instantly available via AI tools such as ChatGPT or, more recently, Google's AI Overviews, without us having to follow links to online sources or publications (Simonetti & Blunt, 2025; Germain, 2025).

According to ChatGPT 4o's digital oracle, consulted on 19 May 2025, "art is a human expression of creativity, imagination and emotion, intended to evoke beauty, emotional impact or reflection. Over time, definitions of art have shifted. Classical traditions often emphasised beauty and skill; modern and contemporary views embrace disruption, criticism and conceptual depth. Today, almost anything can be art – if it is intended as such and invites engagement." (OpenAI, 2025).

The Black Box of AI

While the rapid generation of this answer in a split second is impressive, and appears to have been achieved by machine intelligence, it is not an original creation but a vague synthesis of existing human knowledge, labour, and expression, without any transparent reference to these sources or context. It is precisely these human processes, labour, rights, and context that are all too often forgotten or ignored in the current, mediatised AI discourse. This makes AI appear magical, almost transcendent, but this illusion of supposedly autonomous machine output conceals a complex chain of human labour: programmers, researchers, journalists, writers, artists, etc. All of them contribute, often invisibly and unpaid, to the computational models that generate these magical images, answers, and texts. This phenomenon is often described as the 'black box' of AI: a system whose internal workings are deliberately or unintentionally closed to public scrutiny, and in which input and output are visible, but the intermediate, often human processes and labour are opaque (Hassija et al., 2024; Pasquale, 2015).

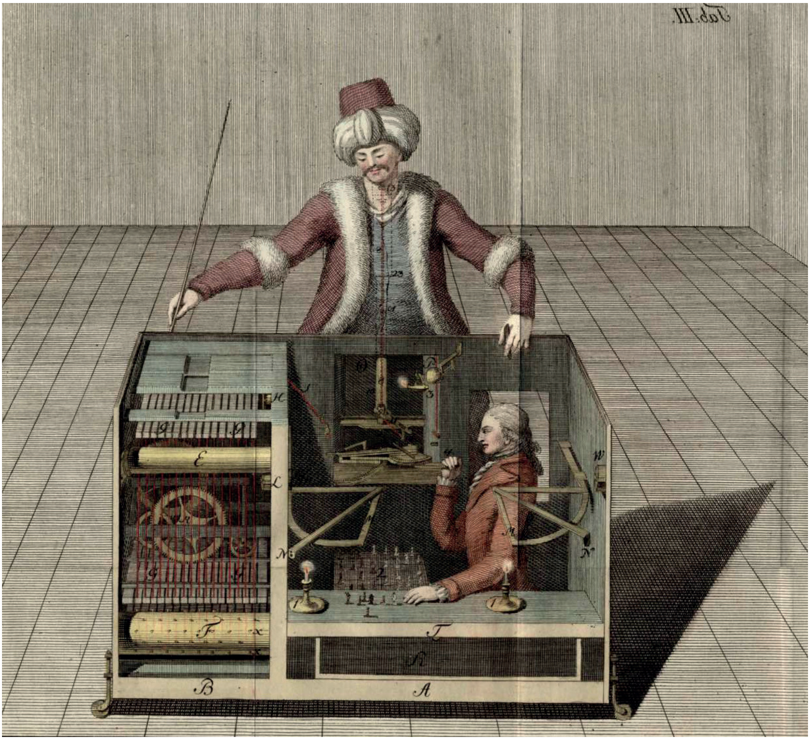

Figure 3: Racknitz, Joseph Friedrich zu. Ueber den Schachspieler. Des Herrn von Kempelen und Dessen Nachbildung. Breitkopf, 1789.

The concept of hidden human labour is illustrated by the historical metaphor of the Mechanical Turk: a chess-playing machine invented by Wolfgang von Kempelen in the 18th century that created the illusion of an autonomous, intelligent machine (Ashford, 2017; Klapper, 2021). This sensational, almost magical invention gave Europe's and later North America's most notable figures the opportunity to play chess against this marvellous machine. One of these figures was reportedly Napoleon Bonaparte, who lost the game. The Mechanical Turk consisted of a large wooden cabinet containing an ingenious mechanical device, topped by a life-size, human-looking figure dressed as a mythical Eastern character. This robot-like figure could move the chess pieces mechanically with his left hand, and when he had checkmated his opponent, a mechanical cry of "Échec! Échec!" was said to sound from the machine (Ashford, 2017). Magic!

It took more than fifty years before this hoax was discovered, and it became clear that there was not only an ingenious mechanism inside the wooden cabinet, but also a human chess master who operated the mechanics and movements of the robot from a secret compartment (see Figure 3). Edgar Allan Poe wrote about the Mechanical Turk in his *Maelzel's chess player* (1836), in which he demonstrated that fraud was involved, thus debunking the 18th-century technological myth of an intelligent machine and exposing the human labour that lay hidden inside the cabinet (Ashford, 2017; Stephens, 2023; Zylinska, 2022).

The story of this 19th-century chess machine is still relevant today in our attempts to understand and make GenAI transparent.

After all, technology is not neutral but the product of hidden structures, interests, and labour (Hassija et al., 2024). In 2001, Jeff Bezos founded a global digital marketplace called Amazon Mechanical Turk, where microtasks such as identifying objects in images, answering questions, moderating content, or writing short texts were outsourced to contractors. These were hired as a cheaper alternative to automated systems, because human labour was actually cheaper for performing certain tasks than programming algorithms. Each microtask was paid only a few cents (Ashford, 2017; Stephens, 2023; Zylinska, 2022).

The hidden human labour of these Mechanical Turks, together with our collective online content, forms the backbone of current AI models, ranging from computer vision to LLMs. They are trained on existing, publicly available online data created by human labour and collected through extractive processes, such as web scraping, which rightly raises serious questions about copyright infringement (Centivany, 2024; Zylinska, 2022). In 2021, British researcher and artist Johanna Zylinska created an interesting work about this invisible labour within AI entitled *View from the window*. Via the Amazon MTurk platform, she asked 100 Mechanical Turks to take a photo of the view from their window. Zylinska's work, in which she collected all the images taken by MTurk contractors in a book, referred not only to the hidden human labour but also to the first photograph ever taken, in 1826 by Joseph Nicéphore Niépce: *View from the window at Le Gras*.

Repeating the Past

Just like GenAI today, the invention of the photographic process was disruptive to the arts and society. For the first time, a machine could not only accurately capture a realistic image of reality but also reproduce it mechanically. Suddenly, painting seemed obsolete. After all, was it still necessary to depict life through paintings when a device could do so more accurately and mechanically? Could a photographic image be considered art? Some artists tried to imitate photographic realism, but others, such as Cézanne, or later Van Gogh, sought something else: emotion, perception, and interpretation (Heilbrun, 2009; Sweet, 2021). In response to the photographic process, impressionism emerged, depicting the world in a new way, followed later by expressionism, which focused more on portraying the artist's inner world. At the same time, artists began to recognise that photography was not merely a mechanical or automated process but also a medium for artistic expression, shaped and operated by humans. The invention of photography opened a new window onto the world, and artists seized this opportunity to explore alternative ways of representing a rapidly changing society, fundamentally transforming the arts and liberating them from existing boundaries and paradigms.

Co-Creation and Artistic Research

Today, GenAI is opening a new window within the arts to explore our rapidly changing society. Just as with the invention of photography in the 19th century, art will enter a new phase. A phase that will be created by the use and exploration of new algorithmic possibilities. As with any disruptive technological shift, we are on the brink of new possibilities and opportunities, but also challenges. We are still clinging to existing definitions and aesthetics, but we will soon discover new forms of art and image-making. Contemporary artists such as Trevor Paglen, Anna Ridler, David Claerbout, Dries Depoorter, and Tom Van der Borght are already exploring these new possibilities in their work, using new technology not to imitate the existing but to create something new that was not possible before. In doing so, they work with their own trained datasets or their own artistic methods to create their artwork.

Boris Eldagsen's image was indeed not a photographic image in the traditional sense. According to Liv Hausken, such images represent a new form of photorealism, namely a compositional construction that bears visual resemblance to a photographic image but is essentially something else. She therefore advocates a nuanced, rather than polarised, debate about images generated with GenAI (Hausken, 2024). The image *Pseudomnesia: The electrician* was not captured with a camera, nor did it use light-sensitive film or a digital sensor. Nor was it an autonomously generated work of art, produced solely by GenAI. Eldagsen's image was the result of a co-creation between disruptive technology based on online image archives and human intervention. Eldagsen trained the GenAI model for several weeks before creating the award-winning image with the technology. He describes his working method as 'promptography', a term that refers to the textual input used to generate the image. In an interview with *The Guardian* in 2023, Eldagsen said that he does not see AI as a threat to his creativity but rather as something that liberates him as an artist from existing boundaries (Hausken, 2024; Williams, 2023).

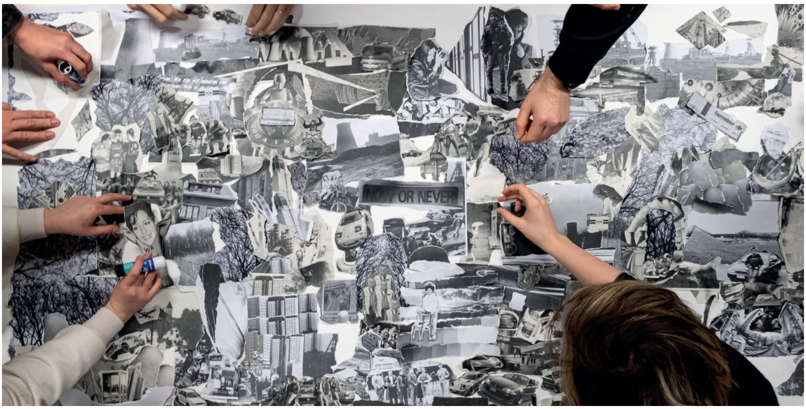

Figure 4: Transisthor (process), Lennert Berx.

In my artistic research practice, using GenAI and LLMs has opened up new possibilities as well. For the artistic and participatory art project 'Transisthor' (see Figures 4 and 5), Dr Niek Kosten and I collaborated with young people from Genk and used GenAI to create a permanent installation. This 6m-long photo collage, consisting of more than a hundred composite images, generates sustainable solar energy for the local community centre while serving as a visual exploration of future imaginaries. Starting from a historical local photo archive spanning more than three decades and documenting the former mining community, we combined this material with newly generated images. These images, created using GenAI and situated prompting, depict speculative future scenarios. Through co-creation with the young people and GenAI, we were able to visualise both the past and the desired futures.

Figure 5: Transisthor (final artwork)

A second example I would like to highlight in this essay is the ongoing research project ‘Changing Gaze, Changing Perception’, in which I collaborate with Baldwin Van Gorp, Niels Hendriks, and Lisa Bormans.

Together, we investigate and visually represent the experiences of people living with dementia (Bormans & Van Gorp, 2025). In an initial experiment using DALL·E2, we prompted the model to generate images of people living with dementia (PLWD) without providing any additional textual input. Despite offering 60 different prompts, we repeatedly encountered the same stereotypical imagery: figures holding their heads in the so-called head-clutcher pose, a motif frequently found in media portrayals of dementia (Van Gorp & Vercruysse, 2012). Moreover, the images lacked diversity.

Towards Transparent and Accountable AI

Such highly biased outputs, if used uncritically, risk reinforcing harmful biases and misconceptions. As creators and researchers, we have a responsibility to challenge these patterns and avoid reproducing them. In response to this problem, we turned to a participatory method: co-creating images with PLWD and GenAI. Through live prompting and iterative feedback sessions, participants were able to specify what needed to change in the images. In one case, it took 25 iterations to arrive at an image with which a particular participant could fully identify. He described the final image (see Figure 6) as an authentic representation of his joy and passion for art, created only weeks before his request for euthanasia was granted. *This was the image he wished to be remembered by*. These images, created together with PLWD and GenAI, are of great value and help us to combat prejudice and stigma and to work towards new, inclusive forms of art and communication, but also affirm the agency and individuality of PLWD. 'Changing Gaze, Changing Perception' demonstrates how participatory GenAI methods can both expose algorithmic biases and facilitate the creation of inclusive, dignified representations. The project underscores the necessity of critical awareness and affirms the indispensable role of human agency and labour in GenAI-driven image production.

Figure 6: 'Changing Gaze, Changing Perception' (image generated with DALL·E2)

It remains crucial that we, within the arts, continue to ask critical questions and actively shape the futures we desire, futures in which we learn to co-work with GenAI. Contemporary art, I believe, has a vital role to play: to make visible what remains hidden, and to critically interrogate the socio-ecological impact of algorithms. In doing so, we must remain vigilant about our human creations, our labour, and our rights. We must not surrender to technological illusions or capitalist imperatives. Instead, we must open the black box, making space for co-creation, transparency, and shared responsibility for equitable, bias-aware, and sustainable futures.

References

Ashford, D. (2017). The mechanical Turk: Enduring misapprehensions concerning artificial intelligence. *The Cambridge Quarterly, 46*(2), 119–139. https://doi.org/10.1093/camqtly/bfx005

Bormans, L., & Van Gorp, B. (2025). How framing shapes dementia research and design: Subliminal influences and considerations. In R. M. Branco, D. Figueiredo, A. De Coen, & C. Craig (Eds.), *Dementia Lab 2025: Ethics + aesthetics in design* (Vol. 5, pp. 43–51). Springer. https://doi.org/10.1007/978-3-031-84901-5

Caramiaux, B., Crawford, K., Liao, Q. V., Ramos, G., & Williams, J. (2025). Generative AI and creative work: Narratives, values, and impacts. arXiv. http://arxiv.org/abs/2502.03940

Centivany, A. (2024). Mining, scraping, training, generating: Copyright implications of generative AI. *Proceedings of the Association for Information Science and Technology, 61*(1), 68–79. https://doi.org/10.1002/pra2.1009

Germain, T. (2025, June 13). *Is Google about to destroy the web?* BBC. https://www.bbc.com/future/article/20250611-ai-mode-is-google-about-to-change-the-internet-forever

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2024). Interpreting black-box models: A review on explainable artificial intelligence. *Cognitive Computation, 16*(1), 45–74. https://doi.org/10.1007/s12559-023-10179-8

Hausken, L. (2024). Photorealism versus photography: AI-generated depiction in the age of visual disinformation. *Journal of Aesthetics and Culture, 16*(1), Article 2340787. https://doi.org/10.1080/20004214.2024.2340787

Heilbrun, F. (2009). Impressionism and photography. *History of Photography, 33*(1), 18–25. https://doi.org/10.1080/03087290802582889

Klapper, S. (2021). Mechanical Turk jurisprudence. *Brooklyn Law Review, 86*(2), 291–324. https://heinonline.org/HOL/License

Knight, W. (2023, January 12). *Where the AI art boom came from—and where it's going*. Wired. https://www.wired.com/gallery/wherethe-ai-art-boom-came-from-and-where-its-going/

OpenAI (2025). *ChatGPT* (July 2025 version) [Large language model]. https://chat.openai.com/

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Poe, E. A. (1836). Maelzel's chess player. *Southern Literary Messenger*, April, 318–326.

Simonetti, I., & Blunt, K. (2025, June 10). *Google's new AI tools are crushing news sites*. The Wall Street Journal. https://www.wsj.com/tech/ai/google-ai-news-publishers-7e687141

Stephens, E. (2023). The mechanical Turk: A short history of 'artificial artificial intelligence.' *Cultural Studies, 37*(1), 65–87. https://doi.org/10.1080/09502386.2022.2042580

Sweet, D. (2021). Before and after photography. *Journal of Contemporary Painting, 7*(1–2), 163–175. https://doi.org/10.1386/jcp_00029_1

Van Gorp, B., & Vercruysse, T. (2012). Frames and counter-frames giving meaning to dementia: A framing analysis of media content. *Social Science & Medicine, 74*(8), 1274–1281. https://doi.org/10.1016/j.socscimed.2011.12.045

Williams, Z. (2023, April 19). *'AI isn't a threat' – Boris Eldagsen, whose fake photo duped the Sony judges, hits back*. The Guardian. https://www.theguardian.com/artanddesign/2023/apr/18/ai-threat-boris-eldagsen-fake-photo-duped-sony-judges-hits-back

Zylinska, J. (2022). Undigital photography: Image-making beyond computation and AI. In T. Dvorák & J. Parikka (Eds.), *Photography off the scale* (pp. 231–252). Edinburgh University Press. https://doi.org/10.1515/9781474478847-015

Governing AI at Work: Anticipating Technological Trends and the Role of Social Dialogue

Aída Ponce Del Castillo

Introduction

The governance of AI requires more than simply reacting to technological change: it demands a forward-looking understanding of where AI is heading, how it is applied, and what its consequences are. Yet anticipating the trajectory of AI presents a number of challenges. Firstly, the technology is evolving at a rapid and often unpredictable pace. Secondly, there is frequently a disjuncture between the intentions of developers and the real-world applications of AI systems. An algorithm designed for one purpose may, in practice, be used in ways that were never foreseen. These challenges become particularly salient in the world of work. As AI systems are introduced into workplaces, they reshape organisational structures, alter job roles, and redefine labour relations. Their impact extends beyond technical efficiency, influencing how decisions are made, how power is distributed, and how workers experience autonomy and control. In such a context, governing AI effectively calls for more than technical oversight or regulatory compliance. It requires an anticipatory approach, one that recognises uncertainty not as a problem to eliminate but as a reality to navigate. This, in turn, calls for inclusive and sustained forms of social dialogue. If foresight is to be meaningful, it must be embedded in democratic processes that bring together policymakers, employers, workers, and other stakeholders to shape how AI is developed and deployed. Anticipation, then, is not a one-off exercise but an ongoing and collective practice.

Technological Trends to Watch

The deployment of AI in the workplace is unfolding against the backdrop of several interrelated technological trends that merit close attention (FTSG, 2025). Among these, three stand out for their transformative potential: 1) the emergence of autonomous AI agents, 2) the integration of affective computing, and 3) the convergence of advanced technologies. In parallel, the growing number of AI-related incidents highlights the pressing need for robust governance and institutional preparedness.

Firstly, the development of AI agents marks a shift from the automation of discrete tasks to the automation of goal-directed agency. These systems are designed to pursue objectives over time without having their actions exhaustively preprogrammed (Shavit et al., 2023).

In the workplace, such agents can take over certain functions, operate semi-independently, and interact with humans in real time. While this opens up opportunities for collaboration between human and machine, it also introduces new forms of liability, surveillance, and algorithmic decision-making. The delegation of agency to adaptable and learning systems inevitably raises critical questions about human autonomy, responsibility, and control.

Secondly, affective computing, which encompasses emotion recognition technologies, sentiment analysis, and behavioural prediction, has made its way into recruitment, employee monitoring, and performance management (Moore et al., 2024). These technologies promise to enhance productivity or well-being by tailoring responses to emotional cues, but they also risk crossing into manipulation and behavioural nudging (Pasquale, 2024). The European Union's AI Act (Article 5(1)(f)) explicitly prohibits emotion inference in workplace and educational contexts, citing fundamental rights concerns. However, exceptions, particularly for medical or safety purposes, remain, potentially creating loopholes. These carve-outs must be scrutinised, and their use actively monitored by worker representatives and regulators alike.

Thirdly, the convergence of advanced technologies is reshaping the landscape in less predictable but no less consequential ways. AI is increasingly embedded within broader assemblages that include neurotechnologies, nanotechnologies, quantum computing, and synthetic biology (WEF, 2025). In the workplace, such convergence can result in novel applications but also unanticipated risks, especially where complex systems interact with human cognition, biology, or social dynamics. Understanding these layered interactions will require interdisciplinary foresight and adaptive governance mechanisms.

Finally, AI-related incidents are on the rise, underscoring the real-world risks posed by the integration of AI into organisational life. Since 2020, reports of harms linked to AI systems have increased significantly, with 2024 marking a particularly steep uptick (AIID, 2025; HAI, 2025). In workplace contexts, many incidents concern breaches of confidentiality: AI-powered meeting assistants inadvertently capture and transmit sensitive information, GenAI tools are deployed without adequate training, and interactions with systems like ChatGPT expose companies to reputational, legal, and data security vulnerabilities (OECD, 2025). Taken together, these trends reflect a complex and rapidly evolving terrain. To navigate it effectively, workers and their representatives must be equipped with not only technical understanding but also institutional spaces for dialogue, deliberation, and democratic oversight.

The EU's AI Act: Informing and Involving Workers

In the context of workplace automation, the European Union's AI Act introduces a legal framework that unfortunately does not directly address the relationship between AI systems and those most affected by them: workers. At work, employers are either providers or deployers of AI systems. Employees are typically their primary users and consequently, the group most exposed to the risks and consequences of such technologies.

However, several provisions in the AI Act do seek to establish minimum safeguards for transparency, oversight, and worker involvement. Article 26(7) obliges deployers to inform workers and their representatives prior to the use of any high-risk AI system. This obligation is reinforced by Article 26(11), which extends the requirement to situations where AI systems contribute to decisions that directly affect individuals. Complementing these is Article 50(1), which introduces a transparency clause for general-purpose AI systems: individuals must be clearly informed when they are interacting with an AI system.

The AI Act also touches on the crucial issue of AI literacy. According to Article 4, both providers and deployers are required to ensure, to the best of their ability, that those who operate or engage with AI systems have the necessary notions to make informed decisions regarding such systems. This includes taking into account users' technical knowledge, experience, education, and the specific contexts in which the systems are applied. This requirement extends to not only staff but also other individuals acting on behalf of the employer, and it explicitly considers the groups or individuals on whom the AI systems are used.

Together, these provisions establish a baseline of worker information, consultation, and education on which more robust and future-oriented forms of social dialogue can be built. Crucially, the AI Act does not limit the possibility for more ambitious protections. Article 2(11) clarifies that national or European laws, as well as collective agreements, may introduce stronger safeguards in favour of workers. This opens the door for sectoral, national, or company-level negotiation and regulation, and reaffirms the role of collective bargaining in shaping the governance of AI at work.

Finally, the AI Act's Article 14 underlines the principle of human oversight. High-risk AI systems must be designed and implemented in ways that allow effective supervision by natural persons throughout their period of operation. This design principle is key to ensuring accountability and upholding workers' rights in increasingly automated environments. In sum, the EU's AI Act provides an essential, though minimal, legal scaffolding for worker protection and engagement in the age of AI. The challenge now lies in translating these principles into meaningful practices through sustained social dialogue and institutional innovation.

Future Ways Forward

The technological developments outlined above, such as affective computing, agentic AI, and technological convergence, are poised to reshape the nature of work in profound and complex ways. As these systems are progressively introduced in workplace settings, their impact will depend on not only how they are designed and deployed, but also how proactively societies prepare for their consequences. To mitigate potential harms, both organisations and regulators must adopt robust anticipatory practices that identify emerging risks and implement safeguards in a timely manner. Foresight should be treated not as a one-off exercise but as an institutionalised and iterative process, integrated into governance structures at multiple levels.

One concrete step in this direction would be the development of real-time incident reporting and early-warning systems, grounded in existing AI incident databases. Such platforms could provide actionable intelligence, improve accountability, and allow organisations to move from reactive risk management towards preventive oversight. This could complement worker consultation, which should remain the primary process ensuring that those on the front lines of AI deployment are also central to shaping its governance.

Ultimately, the effective regulation of AI in the workplace depends on the active involvement of those most directly affected: workers and their employers. Strengthening social dialogue is not a supplementary measure but a structural requirement for democratic and inclusive AI governance. Only through meaningful participation can we ensure that technological innovation aligns with social justice and the dignity of work.

References

AI Incident Database. (2025). *AIID: AI Incident Database*. https://incidentdatabase.ai/apps/incidents/

Future Today Strategy Group. (2025). *Tech trends report*. https://ftsg.com/trends-download/

Moore, P., Barnard, G., & Thomas, A. (2024). *Data on our minds: Affective computing at work*. Institute for the Future of Work. https://www.ifow.org/publications/data-on-our-minds-affective-computing-at-work

OECD (2023). *OECD employment outlook 2023: Artificial intelligence and the labour market*. OECD Publishing. https://doi.org/10.1787/08785bba-en

OECD (2025). *AI incidents monitor: An evidence base for trustworthy AI*. https://oecd.ai/en/incidents

Pasquale, F. (2024). Affective computing at work: Rationales for regulating emotion attribution and manipulation. In A. Ponce del Castillo (Ed.), *Artificial intelligence, labour and society* (pp. 175–179). ETUI. https://www.etui.org/publications/artificial-intelligence-labour-and-society

Reule, A. (2025). *Artificial Intelligence Index Report 2025* (Chapter 3). Stanford University. https://hai.stanford.edu/assets/files/hai_ai-index-report-2025_chapter3_final.pdf

Shavit, Y., Agarwal, S., Brundage, M., Adler, S., O'Keefe, C., Campbell, R., & Robinson, D. G. (2023). *Practices for governing agentic AI systems* [Research paper]. OpenAI. https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf

Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., & Zhang, W. (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion, 83*, 19–52. https://doi.org/10.1016/j.inffus.2022.03.009

World Economic Forum (2025, June). *Technological convergence report* [Insight report]. https://reports.weforum.org/docs/WEF_Technology_Convergence_Report_2025.pdf

Debate Summary

Mapping AI in the Workplace: Challenges and Opportunities

Zeynep Hamurdan, Sercan Kiyak, Frank Pasquale, Meryem Soyturk, Stefan Mertens, Pepijn Viaene, & Elisabeth Struyf

Introduction

Reflected in the global reception of Klaus Schwab's (2016) book, *The fourth industrial revolution*, the idea that we are living in epochally unstable economic times has gained even more popularity recently. Similar to previous "revolutions," the full impact of this technological change remains difficult to estimate. Thinkers on the pessimist side, like Martin Ford, author of *Rise of the robots* (2015), warn of a "jobless future." Yuval Noah Harari (2017) speaks of a coming "useless class." Sober studies also predict troubling effects of AI on job markets (Cazzaniga et al., 2024). While acknowledging concerns, this essay explores reasons for cautious optimism about the future of AI. It begins by defining AI and giving a sense of its history, then turns to the challenges AI poses in the workplace. Finally, it suggests some ways forward, based on a unique interactive dialogue held at the KU Leuven in May 2025.

The Evolution of AI: From Symbolic Reasoning to Generative Systems

AI is not a single technology but a field of diverse computational systems designed to mimic or extend human cognitive abilities. The evolution of AI has unfolded in distinct paradigms. Early AI (1950s–1980s) relied on symbolic reasoning, using explicitly coded "expert systems" that attempted to mimic human logic through static rules (McCarthy, 1983). While effective in narrow domains, these systems were brittle and quickly became outdated—such as early spam filters that blocked specific keywords but failed as tactics evolved.

By the 2010s, deep learning, a subset of machine learning (ML) using multilayered neural networks, became dominant, fueled by increased computational power, abundant data, and improved hardware (LeCun et al., 2015). A milestone was the ImageNet competition in 2012, where a deep neural network outperformed all others, revealing that scale, not just algorithm design, could drive breakthroughs (Krizhevsky et al., 2012). The evolution of spam detection exemplifies this shift: Modern systems use a combination of deep learning, natural language processing, and ensemble models to achieve exceptional results in many areas—far beyond what symbolic or early ML systems could manage.

The next major leap in AI came with GenAI and LLMs built on transformer architecture (Vaswani et al., 2017). Trained on massive corpora of internet text, text-based models produce fluent and coherent outputs in a conversational style. However, these systems rely on probabilistic pattern prediction, not genuine reasoning or understanding. A troubling political economy also underwrites modern AI advances. The evolution of AI is not solely a story of technical progress; it also reflects commercial incentives, extractive data practices, and shifting political power (Crawford, 2021; Green, 2022). The rise of GenAI depends on not only technological breakthroughs but also vast amounts of underpaid data labeling labor, uncompensated use of creative work, and a surrounding culture of hype that makes critical scrutiny and resistance increasingly difficult (Bender & Hanna, 2025).

Today, most AI systems fall into two broad categories: predictive AI, which analyzes past data to forecast future outcomes, and generative AI, which creates new content such as text, images, or code based on learned patterns (Narayanan & Kapoor, 2024). Though often grouped together, these types differ significantly in their goals, applications, and risks: predictive models shape decisions, while generative models reshape communication and expression. Each has already shaped workplaces, and will continue to do so.

The Challenge of AI in the Workplace

The implementation of AI in the workplace fuels public uncertainty about what AI truly encompasses (Wulgaert, 2024). Definitions vary, but the Algemeen Christelijk Vakverbond union (ACV) offers two prominent ones: AI as "all theories and techniques implemented to create machines capable of simulating human intelligence" (Van Thorre, 2022, p. 10), and as "the simulation of human intelligence by technology, primarily by machines such as computer systems" (Maes et al., 2022, p. 4).

Trusting AI on the workfloor begins with transparent integration. ACV stresses that the question is not whether AI should be used but how it can enrich jobs and empower workers through structured social dialogue (Maes et al., 2022). Belgium's Collective Labor Agreement No. 39 offers a legal foothold, obligating employers to notify employee representatives three months prior to implementing impactful technologies (Van Thorre, 2022). However, rapid AI adoption threatens the practicality of this timeframe, suggesting the need for updated regulatory mechanisms.

The moral dilemma lies in AI's lack of moral compass. Van Thorre (2022) insists that while algorithms can assist, they must not replace important human decisions. This concern points to a tension between AI's perceived autonomy and the irreplaceable value of human judgment. Here, UNICEF's definition offers clarity: AI comprises systems that, guided by human-defined objectives, make decisions and adapt behaviorally to their environment (UNICEF, 2021, p. 16). Holmes and Tuomi (2022) commend this view for its inclusion of both data-driven and symbolic systems, and its grounding in human accountability.

ACV identifies three core features of AI—autonomy, self-learning, and reliance on vast data (Maes et al., 2022). These capacities enable AI to perform complex tasks but also create risks such as invasive surveillance, erosion of privacy, and loss of human agency. While AI can reduce bias and enhance flexibility, it can also intensify managerial control and dehumanize work. To establish trust in AI, ACV proposes three safety criteria: legality, ensuring AI respects labor laws; ethical integrity, requiring AI to uphold human values and privacy; and robustness, mandating resistance to errors and breaches (Maes et al., 2022, p. 5). Furthermore, six negotiation principles are suggested: identifying AI's purpose, assessing impact, maintaining human oversight, ensuring transparency, pursuing legitimate aims, and investing in employee development (Maes et al., 2022, pp. 8–9).

Maintaining Human (and Humane) Control of AI in the Workplace

The goals of Maes et al. are valid, and will be pursued at an increasingly granular level. AI is currently at a stage where it assists workers rather than fully replaces them. In many sectors, it has begun to support workers by handling a variety of tasks that are repetitive, time-consuming, or require large amounts of data. AI tools have started to save time and boost productivity among many different sectors, and it is already changing the way we work, but not always in a positive way. Many worry that as AI helps us become more efficient and cost-effective, companies might hire fewer people, pushing unemployment rates higher and widening the gap between the rich and the poor. Additionally, large businesses and tech firms may benefit the most, as they have the funding and space to experiment and adopt more effectively than smaller companies, which could be left behind. This is the reason why there are growing calls for better regulations, clear policies, and social protections that ensure that these developments benefit the public fairly.

These concerns were especially discussed during the May 26 session, "AI and Human Employment," at the University of Leuven's 600th Anniversary AI Dialogues. Speakers from fields such as labor studies, law, industrial relations, and the arts examined the impact of AI on different types of work. One participant explained that AI has been used to adjust the speed of the production line based on the number of workers present. He also mentioned that digital tools now monitor nearly everything, even when employees travel or visit clients. This fuels growing concerns about the constant surveillance that comes with the increased use of these tools, a discussion that needs to be addressed: Where should we draw the line with tracking at work?

Another speaker presented a realistic view of the efficiency discussion, noting that AI is not often used to reduce workloads; instead, it may increase the pressure to deliver results faster. There was also discussion around the value of human knowledge and expertise. One speaker questioned what it says about our society if institutions like the European Commission move toward replacing highly skilled translators with AI tools. She asked: “What kind of society are we building if we say people who can translate in 10 or 12 languages are no longer needed?” This was a very valid example of how AI adoption could send the wrong signal about what, and who, we value. Consider, for example, how endangered translation expertise could eventually erode social capacities to assess the accuracy of machine translation.

Conclusion

While the top executives of AI companies predict that AI will take over jobs in the near future (Abril & O’Donovan, 2025), AI has not yet caused mass layoffs across a broad range of industries. Scholars like Erik Brynjolfsson (2021) and Frank Pasquale (2020) advocate for the development of human-centric AI and support “augmentation” rather than the replacement of human work. As was the case during previous industrial revolutions, technology will eventually displace certain jobs while creating new opportunities. The lump of labor fallacy (whereby automation of a task automatically reduces employment) is just as fallacious now as ever.

The argument for creating human-centric AI is based on the idea that humans should not be viewed as passive actors (Johnson & Verdicchio, 2017; Brynjolfsson, 2021). People are shaping new technologies. There are no inevitable effects of AI. Hiring procedures can ensure more humane and explainable use of AI in hiring. Much of the existing workforce can be retrained to increase the value of their labor and prevent mass layoffs. Efficiency gains via AI should translate into improvements in the quality of working life, not its degradation. The thinkers and ideas discussed above provide roadmaps to achieve all these ends.

References

Abril, D., & O'Donovan, C. (2025, June 20). Bosses want you to know AI is coming for your job: Top CEOs say AI will shrink their workforces — is it posturing or prescience? *The Washington Post*. https://www.washingtonpost.com/business/2025/06/20/ai-ceos-predict-kill-jobs/

Bender, E. M., & Hanna, A. (2025). *The AI con: How to fight Big Tech's hype and create the future we want*. Harper.

Brynjolfsson, E. (2021). The Turing trap: The promise and peril of human-like artificial intelligence. *Daedalus*, 151(2), 272–287. https://doi.org/10.1162/daed_a_01915

Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A. J., Pizzinelli, C., Rockall, E., & Tavares, M. M. (2024). *Gen-AI: Artificial intelligence and the future of work*. International Monetary Fund.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Ford, M. (2015). *Rise of the robots: technology and the threat of a jobless future*. Basic Books, a member of the Perseus Books Group.

Green, B. (2022). *The smart enough city: Putting technology in its place to reclaim our urban future*. MIT Press.

Harari, Y. N. (2017). *Homo Deus: A brief history of tomorrow*. Vintage Books.

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education, 57*, 542–570. https://doi.org/10.1111/ejed.12533

Johnson, D. G., & Verdicchio, M. (2017). AI anxiety. *Journal of the Association for Information Science and Technology,* 68(9), 2267–2270. https://doi.org/10.1002/asi.23867

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25,* 1097–1105.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Maes, E., Van Thorre, M., Gryp, S., & Van Aken, T. (2022). *Artificiële intelligentie op de werkvloer: Gids voor militanten* [Artificial intelligence on the workfloor: Guide for militants]. ACV. https://www.hetacv.be/militant/themas/nieuwe-technologieën/studiedag-ai-op-de-werkvloer

McCarthy, J. (1983). AI: A personal view. In R. Ringle (Ed.), *Toward the year 2000: Work in progress* (pp. 97–102). Ablex Publishing.

Narayanan, A., & Kapoor, R. (2024). *AI Snake Oil*. https://aisnakeoil.substack.com

Pasquale, F. (2020). *New laws of robotics: Defending human expertise in the age of AI*. Harvard University Press.

Pugh, A. J. (2024). *The last human job: The work of connecting in a disconnected world*. Princeton University Press.

Schwab, K. (2015, December 12). The Fourth Industrial Revolution. *Foreign Affairs*. Retrieved June 20, 2025, from https://www.foreignaffairs.com/world/fourth-industrial-revolution

Tubaro, P., & Casilli, A. (2019). *Micro-work, artificial intelligence and the automation delusion* (European Data Protection Supervisor Reports). European Data Protection Supervisor. https://edps.europa.eu/data-protection/our-work/publications/reports/microwork-artificial-intelligence-and-automation_en

UNICEF (2021). *Policy guidance on AI for children*. Author. UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf

Van Thorre, M. (2022). Artificiële intelligentie op het werk? Niet zonder sociaal overleg! [Artificial intelligence at work? Not without social concertation!]. In P. Van Looveren (Ed.), *Vakbeweging 965: Militantenmagazine van het ACV* (pp. 9–15). https://www.hetacv.be/militant/themas/nieuwe-technologie%C3%ABn/studiedag-ai-op-de-werkvloer

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Wulgaert, R. (2024). *AI in de klas. Praktische gids voor onderwijsprofessionals* [AI in the classroom. Practical guide for education professionals]. Owl Press.

Theme 2
AI & Environmental: Sustainability

The Role of Environmental Sustainability in the AI Act

Peggy Valcke

Environmental Sustainability in AI Ethics and Policy Discussions

In his keynote 'AI in a Digital Utopia' at CiTiP's 30th anniversary conference in Leuven in October 2019, philosopher and Yale University Professor Luciano Floridi argued for a new marriage between 'the Green' and 'the Blue' – the Green of all our habitats and the Blue of all our digital technologies. He further developed this idea of a new marriage between digital technology and environmentalism, and how it can support and develop a better society and a healthier biosphere, in his book *The ethics of artificial intelligence*, published in summer 2023. I had the pleasure of working with Luciano on AI4 People's Ethical Framework for a Good AI Society, which we presented to the European Parliament in November 2018 (Floridi et al., 2018). The framework translated the four core principles commonly used in bioethics (beneficence, non-maleficence, autonomy, and justice) to the ethical challenges posed by AI, and advocated for adding a fifth one, explicability, understood as incorporating both intelligibility and accountability. We considered "sustaining the planet" as a core element of the principle of beneficence, next to promoting well-being and preserving dignity.

In its Ethical Guidelines for Trustworthy AI of April 2019, the European Commission's Independent High-Level Expert Group on

AI equally highlighted sustainability and the ecological responsibility of AI systems. It included 'societal and environmental well-being' as one of the seven key requirements for the development of trustworthy AI,[2] and advocated for critical examination of the resource usage and energy consumption of AI systems during training and for measures securing the environmental friendliness of the entire supply chain.

The year after, in February 2020, the European Commission adopted its White Paper on AI. It outlined the vision for a future regulatory framework, reaffirming sustainable development as a goal. Attention was being paid to the environmental impact of AI systems and the significant role that digital technologies such as AI can play in achieving the Sustainable Development Goals. Echoing the Ethical Guidelines for Trustworthy AI, the White Paper stated that "[g] iven the increasing importance of AI, the environmental impact of AI systems needs to be duly considered throughout their lifecycle and across the entire supply chain, e.g. as regards resource usage for the training of algorithms and the storage of data" (European Commission, 2020).

Also in its 2022 Strategic Foresight Report – titled 'Twinning the green and digital transitions in the new geopolitical context' – the European Commission highlights the need to make digital technologies more energy-efficient (European Commission, 2022). Now, what is left of all these good intentions in the AI Act (European Union, 2024), the first comprehensive binding regulation on AI in the EU…?

---

2 Ethics Guidelines for Trustworthy AI (2019, p. 19): "AI systems promise to help tackle some of the most pressing societal concerns, yet it must be ensured that this occurs in the **most environmentally friendly way possible**. The system's development, deployment and use process, as well as its entire supply chain, should be assessed in this regard, e.g. via a critical examination of the resource usage and energy consumption during training, opting for less harmful choices. Measures securing the environmental friendliness of AI systems' **entire supply chain** should be encouraged."

The AI Act

The purpose of the AI Act is to improve how the internal market functions and promote the uptake of human-centric and trustworthy AI, while ensuring a high level of protection of health, safety, and fundamental rights. It introduces a risk-based regulatory framework for AI, classifying systems into four categories: unacceptable, high, limited, and minimal risk. The AI Act combines two classic traditions of EU law, namely product safety and fundamental rights protection. At its core are safety requirements that companies must meet before placing a high-risk AI system on the EU market (for further details and a critical assessment, see also: Almada & Petit, 2025; Smuha & Yeung, 2025). Through these safety requirements, as well as the prohibitions of certain uses of AI and the transparency requirements for particular AI systems, the EU legislator aims to reconcile core EU values, such as fundamental rights, democracy, and rule of law, with economic goals, such as innovation and market access.

Environmental protection and sustainability are explicitly mentioned in the AI Act in five areas:

- as a general value the regulation seeks to uphold (Article 1);
- as an aspect to be included in the standardisation requests that the Commission will issue to European standardisation organisations (Article 40);
- in the context of regulatory sandboxes, allowing for the further processing of personal data to develop, train, and test innovative AI systems with sustainability goals in mind (Article 59);
- as an element to include in codes of conduct that the AI Office and the Member States are expected to encourage and facilitate to foster the voluntary application to AI systems, other than high-risk AI systems, of specific requirements (Article 95); and
- as an aspect to take into account in future evaluations of the AI Act, which the European Commission must do for the first time by August 2028, and every three years thereafter (Article 112).

While these references are to be welcomed, there are certainly missed opportunities as well to anchor ecological aspects more firmly into the text of the AI Act.

The Green Amendments that Weren't

The European Parliament certainly had higher ambitions on that front:

Firstly, it pushed for including 'social and environmental well-being' as a general principle applicable to all AI systems, meaning that AI systems would need to be "developed and used in a sustainable and environmentally friendly manner as well as in a way to benefit all human beings, while monitoring and assessing the long-term impacts on the individual, society and democracy" (proposed Article 4a AI Act; amendment 213) (European Parliament, 2023). AI systems would need to take into account state-of-the art methods and relevant applicable standards to reduce the energy use, resource use, and waste, as well as to increase their energy efficiency and the overall efficiency of the system. To that end, recommendations and guidelines and, eventually, targets for sustainability (KPIs) would need to be developed (proposed recitals 46a and 46b; amendments 82 and 83) (European Parliament, 2023).

Secondly, it suggested that providers of general-purpose AI models (referred to as 'foundation models' in the Parliament's text) in particular be subject to obligations in relation to the reduction of energy use and the increase of energy efficiency (including the obligation to design their models with capabilities enabling the measurement and logging of the consumption of energy and resources) (proposed Article 28b; amendment 399) (European Parliament, 2023).

Thirdly, in the list of implementation guidelines to be adopted by the Commission, it added guidelines on the measurement and logging methods to enable calculations and reporting of the environmental impact of systems, including carbon footprint and energy efficiency (proposed Article 82b; amendment 685) (European Parliament, 2023).

Lastly, it also advocated for specific public funding incentives for environmentally beneficial AI R&D (proposed Article 54a AI Act; amendment 516) (European Parliament, 2023). Parliament also suggested reinforcing the wording for the evaluation reports by the Commission, requiring that those would devote specific attention to "the impact of the Regulation with regards to the resource and energy use, as well as waste production and other environmental impact" (amendment 695) (European Parliament, 2023).

Those amendments were not adopted in the final text, with sustainability largely relegated to preambles or soft law mechanisms (e.g., voluntary codes of conduct).

International Frameworks

The gap between the importance attributed to ecological goals in policy discussions, on the one hand, and the enforceability of legal commitments for states and obligations for undertakings, on the other, is also visible at the international level. I had the privilege of following the discussions on the Council of Europe's Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (Council of Europe, 2024) (also known as the Vilnius Convention, referring to the place where the treaty was opened for signatures on 5 September 2024). From the front row, as head of the Belgian delegation and elected vice-chair of the Ad Hoc Committee on Artificial Intelligence (CAHAI), I could witness how provisions on sustainability were gradually watered down. In both its deliverables, CAHAI[3] – the Committee tasked with analysing the feasibility of a legal instrument on AI at the level of the Council of Europe – explicitly highlighted the need for states to ensure that AI systems are developed and used in a sustainable manner, with full respect for applicable environmental protection standards (CAHAI, 2021; 2020).

Draft versions of the treaty prepared by CAHAI's successor, the Committee on Artificial Intelligence (CAI),[4] consequently included a provision on "preservation of public health and the environment", requiring parties of the treaty to "take the necessary measures to preserve public health and the environment in the context of the application of an artificial intelligence system".[5] In the final version, however, such firm commitment for states was removed and replaced by a mere reference in the preamble to the environment. The reason? Insufficiently clear what such measures would entail, and therefore not possible to expect states to sign up to something so vague…

The Explanatory Report, which elaborates on the Drafters' intention, still contains a non-binding recommendation to duly take into account the need to preserve a healthy and sustainable environment, along with human rights, democracy, and the rule of law. It remains to be seen, though, to what extent sustainability will be taken up, or not, by the Conference of the Parties, the body that will oversee the proper and effective implementation of the Vilnius Convention once it has entered into force.

---

3 https://www.coe.int/en/web/artificial-intelligence/cahai

4 https://www.coe.int/en/web/artificial-intelligence/cai

5 See article 11 in the Revised zero draft from January 2023, available from https://rm.coe.int/cai-2023-01-revised-zero-draft-framework-convention-public/1680aa193f

The Path Forward for Sustainable AI

The attention for, and clear commitments to, environmentally sustainable AI in ethical and policy discussions clearly did not translate well into binding legal texts. Both the AI Act's hard law provisions and the Council of Europe's Vilnius Convention remain cautious and defer many obligations to soft law instruments or future revisions.

The AI Act certainly is a step in the right direction. It establishes a framework for developing harmonised standards aimed at improving AI systems' resource performance, with a particular focus on energy efficiency. But this focus on standards rather than mandatory requirements and the absence of clear enforcement mechanisms is generally considered a missed opportunity to fully tackle AI's environmental impacts (Laranjeira de Pereira, 2024; Warso & Shrishak, 2024; Podder et al., 2024; Garsia, 2025). The AI Act's sandbox system and evaluation clause may provide openings to revisit sustainability in future developments. The Council of Europe's Framework Convention equally shows both political limitations and future potential via its interpretative bodies.

It will require a lot of political courage to integrate sustainability more deeply into future AI governance structures. The prospect does not look so rosy for the moment. Yes, the AI Action Summit in Paris in February 2025 did result in the establishment of a Coalition for Sustainable Artificial Intelligence.[6] Launched by France, the United Nations Environment Programme (UNEP) and the International Telecommunications Union (ITU), and counting 96 partners (including 39 companies, 13 countries, and 6 international organisations), it aims to reduce AI's energy and environmental cost so that it can be widely shared and contribute positively to our energy

6 https://www.elysee.fr/en/sommet-pour-l-action-sur-l-ia

and environmental goals. But when simultaneously hearing President Macron pushing for French nuclear-powered AI in his opening speech at the Summit (with the infamous words "plug, baby, plug") (Caulcutt, 2025), it is hard to resist the thought that the Coalition is mainly there to pay lip service to environmental concerns …

References

Almada, M., & Petit, N. (2025). The EU AI Act: Between the rock of product safety and the hard place of fundamental rights. *Common Market Law Review, 62*(1), 85–120.

CAHAI. (2020, December 17). *Feasibility study on a legal framework on AI design, development and application based on Council of Europe standards* (p. 29). Council of Europe. https://www.coe.int/en/web/artificial-intelligence/cahai

CAHAI. (2021, December 3). *Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law* (p. 7). Council of Europe. https://www.coe.int/en/web/artificial-intelligence/cahai

Caulcutt, C. (2025, February 10). 'Plug, baby, plug': Macron pushes for French nuclear-powered AI. *Politico*. https://www.politico.eu/article/emmanuel-macron-answer-donald-trump-fossil-fuel-drive-artificial-intelligence-ai-action-summit/

Council of Europe (2024). *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (Council of Europe Treaty Series No. 225). https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

European Commission, Independent High-Level Expert Group on Artificial Intelligence (2019). *Ethics guidelines for trustworthy AI* (April 2019). https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

European Commission. (2020, February 19). *White paper on artificial intelligence: A European approach to excellence and trust* (COM(2020) 65 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0065

European Commission (2022, June 29). *Communication to the European Parliament and the Council: 2022 Strategic Foresight Report: Twinning the green and digital transitions in the new geopolitical context* (COM(2022)289 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022DC0289

European Parliament (2023, June 14). Amendments adopted by the European Parliament on the proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html

European Union (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L, 2024/1689, 12 July 2024. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Garsia, S. (2025, 18 February). The environmental costs of AI: a shake-up of the EU's twin transition. *KU Leuven* Law, Ethics & Policy of AI Blog. *https://www.law.kuleuven.be/ai-summer-school/blogpost/Blogposts/the-environmental-costs-of-ai-a-shake-up-of-the-eus-twin-transition & KU Leuven CiTiP Blog. https://www.law.kuleuven.be/citip/blog/the-environmental-costs-of-ai-a-shake-up-of-the-eus-twin-transition/*

Laranjeira de Pereira, J. R. (2024, April 8). The EU AI Act and environmental protection: the case for a missed opportunity. *Heinrich-Böll-Stiftung European Union*. https://eu.boell.org/en/2024/04/08/eu-ai-act-missed-opportunity

Podder, S., Lewis, T., Racz, A., Kobayashi, Y., Menon, N., Lopez, J., Caldeira, V., & Sarro, F. (2024, October 17). The EU AI Act: Insights from the Green AI Committee – Reflections on the first binding regulation on AI globally. *Green Software Foundation articles*. https://greensoftware.foundation/articles/the-eu-ai-act-insights-from-the-green-ai-committee

Smuha, N., & Yeung, K. (2025). The European Union's AI Act: Beyond motherhood and apple pie? In N. Smuha (Ed.), *The Cambridge handbook of the law, ethics and policy of artificial intelligence* (pp. 228–258). Cambridge University Press.

Warso, Z., & Shrishak, K. (2024, May 21). Hope: The AI Act's approach to address the environmental impact of AI. *Tech Policy Press*. https://www.techpolicy.press/hope-the-ai-acts-approach-to-address-the-environmental-impact-of-ai/

Analyzing the Eco-Political Economy of AI to Grasp its Environmental Challenges

Benedetta Brevini

When discussing the risks associated with AI, public narratives often veer toward apocalyptic scenarios imagining sentient chatbots surpassing human intelligence and taking control. In contrast, academic conversations have rightly focused on more grounded concerns: the impact of AI on cognitive abilities, job displacement, algorithmic bias, and the gradual erosion of critical thinking. However, in the last year the public conversation on the harms of AI has shifted dramatically to include, for the first time, the environmental cost of AI. This shift is a result of the unprecedented increase in energy and water consumption driven by the rise of GenAI, making it nearly impossible for western Digital Lords (Brevini, 2020a) to continue avoiding openly discussing this vital issue (Brevini, 2024).

Global data center electricity consumption is on track to more than double by 2030 (IEA, 2025). In Belgium alone, data center usage amounts to 4% of the nation's total electricity use, surpassing the European average of 2%. Its consumption is expected to triple to reach 10% of Belgium's electricity use by 2035 (Brussels Times, 2025). The United States leads with 45% of electricity demand, followed by China (25%) and Europe (15%) (IEA, 2025). Since 2017, global data center electricity consumption has increased at an average rate of 12% per year: this is four times the growth rate of overall electricity consumption, highlighting the extraordinary energy emergency we are facing (IEA, 2025).

This trend raises urgent questions about energy infrastructure, grid resilience, and the environmental sustainability of AI and digital technologies (Brevini, 2021; Falk et al., 2024; Kuntsman, 2020). The sustainability of AI should no longer be considered a peripheral concern. Instead, it must become a central focus of global discussions on the future of technological development, energy policy, social equity, and democratic governance.

While the recent shift in public discourse about AI risks is to be welcomed—particularly for those of us who have long raised concerns about the environmental cost of AI technologies—this contribution argues that it is essential to move beyond narratives that reduce AI's environmental impact to carbon emissions and water usage alone. To fully understand AI's environmental cost, it is essential to assess its global production chain, which involves deeply interconnected processes of extraction, resource depletion, and environmental degradation. From resource extraction to data centers and chip manufacturing, from usage to consumption and disposal, every stage of AI's life cycle is embedded in extractive economies that place significant strain on the planet's resources.

In several publications, I have advocated for the implementation of what I call an Eco-Political Economy of AI: a framework designed to address the complex and interconnected factors shaping AI's environmental impacts (Brevini, 2021; 2022; 2024). This approach entails the examination of three critical segments of the extractive global production and supply chain of AI to account for its environmental costs: a) mining and resource extraction; b) consumption, energy use, and carbon footprints; and c) digital waste.

The production chain of AI is a series of steps that turn raw materials and data into the services we use. It starts with extracting rare minerals and metals, like lithium and cobalt, which are used to build the hardware, such as powerful computer chips. Then, data is collected and fed into these computers, where it is processed by complex algorithms to "train" AI systems, teaching them how to perform tasks like translating, understanding speech, or recognizing faces. Later, the trained AI is embedded into devices, apps, or services. At this point users start interacting, or "consuming" AI. Once these objects reach the end of their life, they need to be disposed of and discarded, turning into e-waste or digital rubbish. Each step, from mining materials to processing data to digital waste, is part of the AI production chain and has a considerable environmental toll.

Assessing the Eco-Political Economy of AI

So, from an eco-political economy of AI framework, it is crucial to consider firstly the extraction of rare minerals and metals, resources that are essential for the fabrication of AI hardware. Lithium, cobalt, copper, nickel, and rare earth elements are all crucial for manufacturing AI-powered devices. The demand for these mineral resources is growing at an alarming rate. The European Commission has projected that the demand for lithium in the EU will increase to 18 times the current level by 2030 and to 60 times the current level by 2050 (European Commission, 2020). How sustainable are the practices for extraction of mineral resources? Besides the political and economic precarity of the supply of these minerals mostly in the Global South (with China being a dominant player in this market), mining operations discharge toxic chemicals and generate vast amounts of hazardous waste, posing serious environmental and health harms (Brevini, 2020b; Najar, 2021; Falk et al., 2024). These practices frequently devastate local and indigenous communities and are reminiscent of colonial exploitation with inhumane working conditions.

Take, for example, the mining of cobalt, often referred to as a "conflict mineral" (Maconachie, 2021). Research into the human rights and health concerns of artisanal cobalt mining reveals alarming findings: Most miners lack safety equipment, leaving them exposed to cobalt dust, which can lead to asthma, respiratory issues, and even a potentially fatal condition known as "hard metal lung disease" (Maconachie, 2021). The extractive nature of this tech-colonialism is evident in the communities that bear the brunt of mining activities, often facing environmental degradation, displacement, and the loss of livelihoods (Birhane, 2020). An example is the salt flats of the Atacama Desert in Chile, one of the biggest reserves of lithium in the world. The environmental impact of lithium mining is extensive (Daroqui, 2022). Large amounts of freshwater, a crucial resource in this dry area, is diverted from living communities to support lithium extraction processes. Copper is another vital resource for digitalization, with 70% of its annual production (19.6 million tons) dedicated to electronic applications (Falk et al., 2024). Affected local and indigenous groups have criticized the absence of regulatory environmental standards for these projects that have disrupted indigenous heritage and displaced indigenous communities.

The second segment to consider has received more media attention, due to the paramount energy demands of data centers: the training of models. As previously discussed, engineering studies have produced substantial evidence highlighting the growing environmental unsustainability of AI. In addition to the high energy demands associated with training AI models, it is now increasingly clear that the process of "inference"—the continuous, real-time interactions between users and AI systems—is also becoming more energy-intensive. This trend is placing even greater pressure on data centers, significantly amplifying their overall energy consumption (Desislavov et al., 2023). Anthony et al. (2020) introduced the "carbontracker" as a novel tool designed for monitoring and predicting the energy consumption and carbon emissions associated with training deep learning models (Desislavov et al., 2023). The "carbontracker" not only enables the generation of carbon impact statements but also provides a unique feature that allows users to halt model training at their discretion if the predicted environmental cost is exceeded. In more recent times, tools such as the "machine learning emissions calculator" (Lacoste et al., 2019) have become increasingly accessible (Luccioni et al., 2023). Beyond the carbon footprint, the water footprint is becoming even more alarming. Recent studies focusing on ChatGPT have highlighted the urgency of recognizing the increasing water demands of complex AI models (George et al., 2023; Heikka, 2023).

Following the launch of GenAI services in 2022, both Microsoft and Google reported unprecedented water consumption increases: Google's data centers consumed 20% more water in 2022 compared to 2021 (Google, 2023), while Microsoft's water usage jumped by 34% during the same timeframe (Microsoft, 2022).

The last, crucial link is disposal. When digital devices are discarded, they become electronic waste, leaving local municipalities with the difficult task of ensuring safe disposal. This task is so onerous that it is frequently offshored, with many countries, primarily in the Global South, becoming digital dumping grounds for more privileged nations. According to the UN's Global E-Waste Monitor (2024) the world's production of electronic waste is growing five times faster than documented e-waste recycling. A record 62 million tons of e-waste was produced in 2022—up 82% from 2010. At this rate we are on track to produce 82 million tons by 2030. Only 1% of the demand for rare earth elements is fulfilled through e-waste recycling. GenAI is especially problematic in this regard, as it is driving faster server innovation, particularly in chip design: The latest AI chips such as the Nvidia are driving e-waste up at unprecedented levels (Kidd, 2024).

Solutions, Debates, Opportunities

Despite a growing number of international agreements, position papers, and national-level guidelines, there is still no comprehensive framework capable of addressing the full complexity of AI's environmental challenges. While environmental harms linked to AI have been widely discussed—particularly within the European Parliament—the recently adopted EU AI Act (EU AI Act, 2024) falls short. Though it follows a risk-based approach, it fails to meaningfully incorporate binding environmental regulations (Falk et al., 2024). Instead, it allows providers to voluntarily adopt codes of conduct, which *may* include sustainability commitments—an inadequate response given the urgency of the issue.

This is especially disappointing considering UNESCO's 2021 Recommendation on the Ethics of Artificial Intelligence, which explicitly states that AI systems should not be deployed if they cause disproportionate harm to the environment (UNESCO, 2021). A more promising development came in March 2024, when the U.S. Securities and Exchange Commission (SEC) introduced new climate disclosure rules, which have since been repealed. These regulations required companies to report their emissions, recognizing them as financially material information for investors (U.S. Securities and Exchange Commission, 2024).

These types of interventions could be coupled with an AI Carbon Footprint Label that not only provides details of the energy used for training, but also clearly lists details about the raw materials used, its overheads, recycling options, and disposal costs. This would help increase public awareness of the environmental implications of adopting a particular AI-powered device. Bringing transparency to the making of energy used in producing, transporting, assembling, and delivering everyday AI technology would empower policymakers to make better-informed decisions and enable the public to make more conscious choices.

As expected, the Trump administration, through its significant influence over trade and industrial policy, has created major obstacles to international efforts aimed at aligning industrial and AI development with sustainability goals. This included the withdrawal from the Paris Agreement, which weakened global climate cooperation, and the rollback of domestic environmental regulations, such as SEC regulations, signaling a continuous disregard for ecological accountability. Additionally, the administration has weakened key institutions such as the Environmental Protection Agency (EPA) and dismissed scientific expertise, making it more difficult to establish global norms around responsible AI development, energy use, and environmental justice. The withdrawal from multilateralism and science-driven policy has created a leadership void on the international stage at a critical moment for regulating the environmental impacts of emerging technologies like AI.

In March 2023, the Intergovernmental Panel on Climate Change (IPCC), composed of the world's foremost climate scientists, released its latest report—dubbed the "Final Warning." It emphasized the urgent need to limit global temperature rise to 1.5°C above pre-industrial levels, beyond which climate damage will become irreversible (IPCC, 2022). To stay within this limit, emissions must be cut by at least 43% by 2030 relative to 2019 levels, and by 60% by 2035. Given the growing environmental impact of AI development, achieving these targets is becoming increasingly challenging. It is hoped that this leadership gap will be filled by renewed European initiatives, spearheading AI development grounded in principles that place ecological justice and human wellbeing at its heart.

References

Anthony, L. F. W., Kanding, B., & Selvan, R. (2020). Carbontracker: Tracking and predicting the carbon footprint of training deep learning models (arXiv:2007.03051). arXiv. https://arxiv.org/abs/2007.03051

Birhane, A. (2020). Algorithmic colonization of Africa. *SCRIPT-Ed, 17*(2), 389–409.

Brevini, B. (2020a). Conclusion (pp. 65 -70). In B. Brevini & P. Swiatek (Eds.), *Amazon: Understanding a global communication giant* (pp. 65 -70). Routledge.

Brevini, B. (2020b). Black boxes, not green: Mythologizing artificial intelligence and omitting the environment. *Big Data & Society, 7*(2). https://doi.org/10.1177/2053951720935141

Brevini, B. (2021). *Is AI good for the planet?* Cambridge: Polity.

Brevini, B. (2024). An eco-political economy of AI: environmental harms and what to do about them. *Artificial intelligence, labour and society*, 75.

Brussels Times (2025). By 2035, data centres could consume 10% of Belgian electricity. https://www.brusselstimes.com/1605432/by-2035-data-centres-could-consume-10-of-belgian-electricity

Daroqui, A. (2022). False solutions to climate change: Lithium extraction at the Atacama Desert of Chile. https://globaljusticeecology.org/false-solutions-to-climate-change-lithium-extraction-at-the-atacama-desert-of-chilefalse-solutions-to-climate-change-lithium-extraction-at-the-atacama-desert-of-chile/

Desislavov, R., Martínez-Plumed, F., & Hernández-Orallo, J. (2023). Trends in AI inference energy consumption: Beyond the performance-vs-parameter laws of deep learning. *Sustainable Computing: Informatics and Systems*, *38*, 100857.

European Commission (2020, September 3). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Critical Raw Materials Resilience: Charting a Path towards Greater Security and Sustainability* (COM (2020) 474 final) [Communication]. EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0474

European Parliament & Council of the European Union (2024). *Regulation (EU) 2024/XXX on artificial intelligence* (EU Artificial Intelligence Act). https://artificialintelligenceact.eu/

Falk, S., van Wynsberghe, A., & Biber-Freudenberger, L. (2024). The attribution problem of a seemingly intangible industry. *Environmental Challenges*, 101003.

George, A. S., George, H., & Gabrio Martin, A. S. (2023). The environmental impact of AI: A case study of water consumption by ChatGPT. *Partners Universal International Innovation Journal, 1*(2), 97–104.
Google (2023). Google 2023 environmental report. https://sustainability.google/reports/google-2023-environmental-report/
Halleux, V. (2024, September). New EU regulatory framework for batteries: Setting sustainability requirements (PE 689.337). European Parliamentary Research Service. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)689337
Heikka, M. (2023). Making an image with generative AI uses as much energy as charging your phone. *MIT Technology Review*. https://www.technologyreview.com/2023/12/01/1084189/making-an-image-with-generative-ai-uses-as-much-energy-as-charging-your-phone/
International Energy Agency. (2025, April 10). *AI is set to drive surging electricity demand from data centres while offering the potential to transform how the energy sector works* [Report]. IEA. https://www.iea.org/reports/energy-and-ai/energy-demand-from-ai
IPCC (2022). The evidence is clear: the time for action is now. We can halve emissions by 2030. https://www.ipcc.ch/2022/04/04/ipcc-ar6-wgiii-pressrelease/
Kidd, M. (2024) Energy and e-waste: The AI tsunamis. https://www.datacenterdynamics.com/en/opinions/energy-and-e-waste-the-ai-tsunamis/
Kuntsman, A. (2020). From digital solutionism to materialist accountability: The urgency of new interventions. *Journal of Environmental Media, 1*(1), 15–19. https://doi.org/10.1386/jem_00007_1
Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the carbon emissions of machine learning (arXiv:1910.09700). arXiv. https://arxiv.org/abs/1910.09700
Luccioni, A. S., Viguier, S., & Ligozat, A. L. (2023). Estimating the carbon footprint of bloom, a 176b parameter language model. *Journal of Machine Learning Research, 24*(253), 1–15.
Maconachie, R. (2021). "We miners die a lot." Appalling conditions and poverty wages: the lives of cobalt miners in the DRC. https://theconversation.com/we-miners-die-a-lot-appalling-conditions-and-poverty-wages-the-lives-of-cobalt-miners-in-the-drc-220986
Microsoft (2022). *Environmental sustainability report 2022* [PDF]. Microsoft Corporation. https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/msc/documents/presentations/CSR/2022-Environmental-Sustainability-Report.pdf
Najar, J. (2021). Not so "green" technology: The complicated legacy of rare earth mining. https://hir.harvard.edu/not-so-green-technology-the-complicated-legacy-of-rare-earth-mining/

Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and policy considerations for deep learning in NLP* (arXiv:1906.02243). arXiv. https://arxiv.org/abs/1906.02243

U.S. Securities and Exchange Commission. (2024). *Rules to enhance and standardize climate-related disclosures for investors* (Release No. 2024-31). https://www.sec.gov/newsroom/press-releases/2024-31

UN Global E-Waste Monitor (2024). *UN global e-waste monitor*. https://ewastemonitor.info/wp-content/uploads/2024/03/GEM_2024_18-03_web_page_per_page_web.pdf

UNESCO (2016a). *Education 2030: Incheon Framework for Action—Implementation of SDG 4*. https://uis.unesco.org/sites/default/files/documents/education-2030-incheon-framework-for-action-implementation-of-sdg4-2016-en_2.pdf

UNESCO (2016b). *Education 2030: Incheon Declaration and Framework for Action—Towards inclusive and equitable quality education and lifelong learning for all*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000245656

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO. https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence

van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics, 1*, 213–218. https://doi.org/10.1007/s43681-021-00043-6

Debate Summary

Acceleration to Accountability: AI's Environmental Cost and Ecological Futures Alongside Technology

Kamile Grusauskaite, Tania Azadi, Bettina Berendt, Katja Biedenkopf, Benedetta Brevini, Ans De Nolf, & Bieke Zaman

Introduction

The year 2024 marked a sobering milestone: it was the hottest year ever recorded on Earth. It was also the year in which two of the world's leading AI companies, Google and Microsoft, failed to meet their own climate targets (Zewe, 2025). This convergence of AI acceleration and environmental degradation has intensified global discussions across public, policy, and academic arenas. Is AI a potential ally in the climate crisis, or a hidden driver of ecological harm? What risks does it pose, and how might its rapid normalization in everyday life be pushing the planet closer to a point of no return?

In response to these pressing questions, an interdisciplinary dialogue unfolded among researchers examining the intersections of AI and climate (understood here in a broad sense, encompassing environmental concerns). Their discussion explored the environmental implications of AI's material infrastructure, the limitations of current governance frameworks, and the need to reimagine the future roles of AI in light of ecological constraints.

This essay brings together the key arguments that emerged and outlines several forward-looking recommendations. It advances the core contention that the environmental cost of AI remains largely obscured, both by industry narratives and regulatory inertia. Moving toward a climate-just future in the age of AI will require confronting these hidden costs, developing robust mechanisms for accountability, and critically envisioning more equitable and sustainable technological futures. In this context, references to AI primarily refer to GenAI, as it is the focus of most current debates.

Materiality over Myth. Reckoning with the Materiality of AI

GenAI refers to a class of AI systems designed to produce novel content such as text, images, audio, video, or code by learning patterns from existing data and generating outputs that resemble human-created artifacts. These systems typically rely on generative models, including Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Transformer-based architectures like GPT, to synthesize new data based on learned distributions (Leslie & Rossi, 2023). According to He et al. (2025), GenAI evolutions span rule-based systems, statistical models, deep learning, and large-scale foundation models that result in algorithms capable of producing content that is perceived as exhibiting human-like creativity and adaptability. Ronge et al. (2025) also emphasize that GenAI is best understood through four core dimensions: (multi-)modality, interaction, flexibility, and productivity, which reflect its functional relevance beyond purely technical definitions.

While the development of AI dates back to the 1950s (Charniak, 2024), the launch of GenAI tools, such as ChatGPT in 2022, has catapulted AI into mainstream use and global consciousness. Within just two months, the chatbot amassed over 100 million users (Hu, 2023), a pace of adoption that is both staggering and revealing. Promoted for its efficiency gains and enhanced access to information, AI has quickly become embedded in everyday life. Yet, as our discussion emphasized, this normalization masks significant environmental costs.

The 2024–2025 Global Risks Perception Survey (GRPS) lists the "adverse outcomes of artificial intelligence" among the top ten global threats. These risks are not only abstract or future-oriented, but also already materializing in the form of massive energy and water demands, and increasing pressure on shared power grids. The infrastructure that supports AI, from data centers to supply chains, has a tangible environmental footprint. Moreover, the burden of AI's resource-intensive development is not evenly distributed. As Brevini (2021) argues, the social and ecological costs fall disproportionately on already marginalized communities, exacerbating existing inequalities and raising urgent questions about environmental justice in the age of AI. Rather than viewing AI as a disembodied force of innovation, we must confront its material underpinnings and the extractive systems that sustain it.

What are the Climate Costs of AI?

GenAI operations require significant computational resources. Data centers, for example, are the physical foundation of GenAI infrastructure, housing the high-performance computing systems and storage networks that enable model training, deployment, and real-time inference (Liu et al., 2025). Recent projections warn that data center energy consumption is accelerating at an unsustainable pace. According to the U.S. Department of Energy (2024), electricity demand from data centers is expected to triple by 2028. Even after training AI, energy demands remain high—each ChatGPT query, for example, consumes about five times more electricity than a standard web search.

The training of OpenAI's GPT-3 alone consumed an estimated 1,287 megawatt hours of electricity, enough to power about 120 average U.S. homes for a year, and generated over 550 tons of carbon dioxide (Zewe, 2025). Beyond electricity, data centers also require vast quantities of water for cooling, with each kilowatt hour of energy consumed demanding approximately two liters of water, further straining local ecosystems. Taken together, the planetary costs of these technologies are growing with their popularity, and the acceleration of its development for "innovation" shows no signs of slowing down.

Yet, despite its well-known costs, AI development, driven largely by powerful private actors and enabled by weak regulation, is rarely accountable to its political, social, and ecological consequences. Companies like Google, Microsoft, Amazon, and others continuously fall short of their own (self-imposed) yearly climate pledges, while Amazon even falsely claimed carbon-negative operations until recently (e.g., Kairos Fellowship, 2025). In some cases, instead of taking responsibility for the environmental costs associated with their technologies, many tech leaders attempt to shift public attention by promoting utopian narratives of a technologically enhanced future beyond Earth. Rather than confronting the ecological consequences of AI development, they deflect critique by "selling" yet another grandiose technological dream, this time one that exploits space instead. Tech leaders like Eric Schmidt now promote visions of moving data centers into space (Maruccia, 2025), an expression of a kind of techno-utopianism that masks their underlying dystopian realities. All the while, the environmental cost of hardware production and energy consumption on Earth continues to mount.

Graphics Processing Unit (GPU) shipments to data centers surged from 2.67 million in 2022 to 3.85 million in 2023, driving up emissions from manufacturing, the use of toxic chemicals in raw material extraction, and environmental damage from global transport (HPCwire, 2024). Meanwhile, the utopian and space-related tech narratives can help depoliticize AI by presenting its development as both inevitable and benevolent.

Such trends underscore the inadequacy of corporate self-regulation and the urgent need for robust, binding accountability mechanisms. More than that, in some cases, the industry continues framing AI as a solution to the climate crisis, ignoring its own resource footprint. Some local and international regulation has tried to govern and mitigate its threat. The European Union's AI Act has been hailed as the first-of-its-kind regulatory framework to govern artificial intelligence (European Commission, 2024). However, as became increasingly clear throughout our discussion, the framework only scratches the surface when it comes to addressing sustainability in AI development. Tellingly, the act mentions sustainability or climate only four times in the entire document (for more details, see Peggy Valcke's essay). The result is a regulatory effort that ultimately falls short of the very values it claims to uphold (Brevini, 2022).

This lack of effective oversight is, however, not entirely surprising. It reflects a broader pattern in European digital policy, often shaped by an "opportunities discourse" that emphasizes the promised benefits and imagined futures of technological innovation while downplaying risks. Such discourse tends to depoliticize the structural dynamics underpinning technological development. When the consequences of AI's development and use are obscured by industry narratives and rendered invisible in even the most pioneering policy frameworks, meaningful accountability for its environmental costs becomes nearly impossible.

The question of *how did we get here*—where powerful technologies are developed with little regard for the planet and communities most affected by their expansion—was core to our debate. What are the underlying logics of the current approach to developing and governing AI? The environmental consequences of AI are not merely the result of technical oversights or inefficiencies; they are rooted in a broader political economy that externalizes ecological harm while consolidating technological and corporate power. The infrastructures that sustain AI, from the extraction of rare earth minerals to the operation of energy-intensive data centers, are deeply embedded in global supply chains that reproduce colonial patterns of exploitation (Brevini, 2021). Minerals are extracted in the Global South (Majority World), while devices are primarily designed for use in the Global North and eventually discarded back in the Global South. This cycle perpetuates a familiar form of environmental injustice, with toxic e-waste piling up in places like Nairobi, far removed from the Western centers of innovation and profit (Bore et al., 2017). Moreover, locations such as Nairobi reveal the often-overlooked human labor involved in training AI systems, with Kenyan workers outsourced by OpenAI paid less than $2 per hour to help make ChatGPT less toxic—an example of the human costs that are frequently obscured (Perrigo, 2023). Reclaiming the political in AI governance, as underscored in our discussion, requires naming these dynamics explicitly. It means placing environmental and social justice at the center of regulatory and technological debates, and building interdisciplinary, democratic frameworks capable of holding AI development genuinely accountable to both people and the planet.

Reckoning, Tracking, Futuring

Where do we go from here? If the development of GenAI continues on its current trajectory, racing ahead of governance, extracting energy and labor from increasingly strained planetary systems, it risks deepening global inequalities and environmental degradation. Yet, the future is not foreclosed. In our discussion, we defined some actionable ways to think about and motivate considerations on sustainability and justice in the development, governance, and use of AI.

Reckoning

As noted, a clear trend in the development and governance of AI is the lack of accountability for its environmental and social effects. The road to automation is paved with global, local, and environmental inequalities that must be considered in any attempt to govern AI. Our session proposed that in reckoning with the environmental costs of AI, we must address two main challenges they create: gaps in public awareness and the cultivation of green digital literacy.

First, we must identify public awareness gaps and understanding of the environmental cost of everyday queries. Adding labels for emissions transparency on AI tools could be one of the accountability mechanisms, originally developed for food or energy efficiency standards that keep the developer (even a little) more accountable and make the user aware.

Second, we must cultivate green digital literacy. This includes initiatives such as those supported by UNESCO, which advocate for integrating sustainability and digital literacy into early education, teaching children to understand the environmental footprint of their online activities. A related effort is the Greening Education

Partnership, which adopts a whole-of-system approach to help countries address the climate crisis by harnessing the transformative power of education. As a collaborative platform, it brings together governments, intergovernmental organizations, civil society, youth, academia, and the private sector to embed climate action across all aspects of education systems. Such initiatives highlight the importance of fostering ecological awareness and accountability from an early age, equipping future generations to critically engage with the digital technologies shaping their world. Additionally, engineering programs should include mandatory "green AI" training to equip future developers with the knowledge and responsibility to consider energy, water, and material use in the systems they design. Green literacy is not a luxury secured to the Global North; it is a critical global necessity in shaping AI for ecological futures.

Tracking

Yet, recognition is only a start. To move from recognition to responsibility, we need to develop tools and frameworks to track the full environmental footprint of AI systems, from data and energy consumption to material extraction, labor, and e-waste.

Recent research shows that the environmental costs of GenAI are not limited to training. Therefore, we should first recontextualize how we measure and track the footprint of AI to include the consumption from interface processes, GPU manufacturing, hardware sourcing, and transport, which all contribute to exceeding carbon and water footprints (Zewe, 2025). This means considering and tracking the entire supply chain from raw material to end-of-life disposal.

One promising approach within the European context is the Energy Efficiency Directive. Its promise lies in its legal mandate to measure energy performance and enforce efficiency standards (European

Commission, 2023). Yet, it remains a potentially powerful but underutilized instrument for assessing and regulating the impact of ICT infrastructure, including AI. Comprehensive and enforceable tracking is essential to take the environmental cost of AI out of its "black box" and move toward meaningful regulation, corporate accountability, and innovation that is sustainable, not detrimental.

Futuring

Yet, tracking and implementation should not hinder meaningful questions about the use of AI in our societies, particularly in the face of the climate crisis. Throughout our debate, our positionality as a group of interdisciplinary female scholars made us think about not only how to govern the seemingly ungovernable, but also what imaginaries and values underlie the current race for AI development. Drawing from traditions of eco-feminist and decolonial thought on technology (Puig de la Bellacasa, 2017; Haraway, 2016; Benjamin, 2019), we explored the question of what it would mean to approach AI with the values of care, interdependence, and justice, rather than acceleration and profit. Such reimagining must first take place in three domains: (1) governance, (2) infrastructures, and (3) education.

In terms of governance, revising existing regulatory frameworks is a top priority. The European AI Act marks a shift from soft law to binding obligations, yet concerns remain about enforcement, international coherence, and its reliance on industry self-assessment in certain areas (see Peggy Valcke's essay in this book). To be able to address the environmental impacts of these technologies, the act must evolve to bind legal framework to place climate at its center. This would involve mandating measurable environmental standards across the AI lifecycle, from development to deployment to energy use and disposal. It also requires advancing complementary efforts like the Council of Europe's CAHAI framework, which is currently under revision and offers an opportunity to embed stronger commitments to environmental and societal well-being in international AI governance.

Infrastructures must also be reconsidered. Today's AI systems are built on extractive, energy-intensive, and hyper-centralized infrastructures. Often, these are owned and dominated by a handful of corporate cloud providers. This concentration of technological power not only deepens inequality but also limits democratic oversight and ecological responsiveness. A viable alternative lies in the development of commons-based infrastructure that is publicly supported and decentralized. Such infrastructures could redistribute computational resources, reduce dependence on megascale data centers, and offer communities greater control over the technologies that shape their lives. Reimagining infrastructure means not only greening the material layers of AI but also transforming who owns and governs the systems underneath. This transformation must place diverse people, not just dominant groups, at the heart of AI governance, ensuring broad and equitable participation in shaping these technologies. Inclusive governance frameworks can help dismantle hierarchies of power and allow marginalized communities to co-create systems that reflect their values, needs, and ways of knowing. Without this dimension, even decentralized infrastructures risk reproducing exclusion under a different banner.

Education is a third and foundational domain where meaningful shifts can occur. Transforming the future of AI for climate should be rooted in rethinking the values that we embed in our systems of knowledge production. Throughout our discussion, we concluded that this starts with challenging dominant narratives that equate "more AI" with "progress," particularly those rooted in scarcity-driven logics that justify expansion because "everyone else is doing it." Instead, we propose a culture shift toward a slower, thoughtful tech that embraces resonance, resilience, reflection, and care as essential. Schools and universities should protect time for critical thinking, collective judgment, and ethical deliberation. The example of KU Leuven's voluntary AI ethics program and the KU Leuven DigiSoc Digital Ethics postgraduate program are promising models, yet, in times of climate crisis everyone must get on board. Therefore, in a world that relies more and more on technology, subjects like sustainability, ethics, and justice should be made core components of education across all disciplinary domains, not electives. In computer science, this is also being demanded by major professional associations like the ACM and implemented in various universities (Denoo et al., 2024), for example at TU Berlin.

Inspired by well-being economics, particularly the French school of thought that places collective care at the center of policy, AI education can still be reframed to prioritize collective well-being over supposed "innovation" and short-term gains. In this light, education is an important infrastructure for building just and livable futures in the digital era.

References

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity Press.

Bore, E. K., Were, S., & Cheruiyot, W. (2017). Determinants of the adoption of technology in e-waste management at the Nairobi Central Business District. *International Journal of Informative Research (IJIR)*, 3(1). http://imperialjournals.com/index.php/IJIR/article/view/3258

Brevini, B. (2021). *Is AI good for the planet?* Polity Press.

Charniak, E. (2024). *AI & I: An intellectual history of artificial intelligence*. MIT Press.

Denoo, M., Dupont, B., & Zaman, B. (2024). Many faces, many names? Ethics in Belgian game development education. *ACM Games: Research and Practice, 2*(2), Article 15, 1–23. https://doi.org/10.1145/3675804

Elsner, M., Atkinson, G., & Zahidi, S. (2025, January 15). *Global risks report 2025* (20th ed.). World Economic Forum. https://www.weforum.org/publications/global-risks-report-2025

European Commission (2024). *AI Act: Shaping Europe's digital future*. Digital Strategy. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European Commission (2023, September 20). *Energy Efficiency Directive (EU 2023/1791): Delivering energy savings across the European Union*. https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficiency-targets-directive-and-rules/energy-efficiency-directive_en

Haraway, D. (2016). *Staying with the trouble: Making kin in the Chthulucene*. Duke University Press.

He, R., Cao, J., & Tan, T. (2025). Generative artificial intelligence: A historical perspective. *National Science Review, 12*(5), nwaf050. https://doi.org/10.1093/nsr/nwaf050

HPCwire (2024, June 10). Nvidia shipped 3.76 million datacenter GPUs in 2023, according to study. *HPCwire*.

Hu, K. (2023, February 1). *ChatGPT sets record for fastest-growing user base – analyst note*. Reuters. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Kairos Fellowship (2025, July 2). *Google's ecofailures: How increased AI and data center expansion are hiding rising emissions* [Report]. No Climate Results Found. Retrieved from https://noclimateresultsfound.com/report/

Leslie, D., & Rossi, F. (2023). ACM TechBrief: Generative artificial intelligence. In *ACM eBooks*. Association for Computing Machinery. https://doi.org/10.1145/3626110

Liu, Y., Du, H., Niyato, D., Kang, J., Xiong, Z., Wen, Y., & Kim, D. I. (2025). Generative AI in data center networking: Fundamentals, perspectives, and case study. *IEEE Network*. Advance online publication. https://doi.org/10.1109/MNET.2025.3563262

Maruccia, A. (2025, May 5). Former Google CEO Eric Schmidt wants to put data centers in space. *TechSpot*. https://www.techspot.com/news/107801-former-google-ceo-eric-schmidt-wants-put-data.html

Perrigo, B. (2023). OpenAI used Kenyan workers on less than $2 per hour to make ChatGPT less toxic. *Time*. https://time.com/6247678/openai-chatgpt-kenya-workers/

Puig de la Bellacasa, M. (2017). *Matters of care: Speculative ethics in more than human worlds*. University of Minnesota Press.

Ronge, R., Maier, M., & Rathgeber, B. (2025). Towards a definition of generative artificial intelligence. *Philosophy & Technology, 38,* Article 31. https://doi.org/10.1007/s13347-025-00863-y

U.S. Department of Energy (2024, December 20). DOE releases new report evaluating increase in electricity demand from data centers. https://www.energy.gov/articles/doe-releases-new-report-evaluating-increase-electricity-demand-data-centers

Zewe, A. (2025, January 17). Explained: Generative AI's environmental impact. *MIT News*. https://news.mit.edu/2025/generative-ai-environmental-impact-0117

Theme 3:
AI & Urban Life

Urban AI and the Smartification of African Cities: Forms, Examples, and Challenges

Katrien Pype

AI is increasingly shaping urban futures across the globe, and Africa is no exception. This essay builds on previous research carried out with an interdisciplinary collective of four female researchers—two from the Global North and two from the Global South—exploring social and ecological justice for urban AI, and developed around Cairo. It led to the Cairo Charter (El Khateeb et al., n.d.). Egypt provides an illuminating case study of urban AI development. Since 2014, the government has launched 14 "fourth-generation" cities equipped with the latest technologies, including AI. These cities are planned in collaboration with various ministries and executed by the New Urban Communities Authority. Our Cairo Charter outlines principles for justice-driven urban AI. We organized designathons, in which Cairo's urban planners, technologists, and citizens brainstormed appropriate AI design for their city. Although tailored to Cairo, the strategies stipulated in the Cairo Charter are applicable to other African cities and cities in the Global South. In addition, I also integrate arguments and findings drawn from my earlier ethnographic research in Kinshasa, capital city of the Democratic Republic of Congo (Pype, 2017; 2021). The tech ecosystems of the two cities are entirely different, thus allowing for a nuanced understanding of what AI can be and do in African cities.

Urban AI

In the Cairo Charter, we defined "urban AI" as follows:

"Urban AI, or urban artificial intelligence, is reshaping cities through embedded hardware and software technologies including natural language processing, facial recognition, machine sensing and processing, and the Internet of Things. These technologies can be found in everything from buildings to vehicles, with sensors scaling from molecules to infrastructure. Urban AI adds a new dimension to the capacities of smart cities, transforming the ways that information is captured, shared, transmitted, and how decisions are made. (decisions by citizens, civil society, urban authorities and so on)." (El Khateeb et al., n.d., p. 4).

Urban AI, therefore, is not only a technological development but also a socio-political and economic phenomenon. Depending on the context in which urban AI is integrated, it can reflect aspirations for modernization, national pride, and improved quality of life.

Two Forms of Urban AI

Urban AI in Africa manifests in two primary forms: "smarter cities" and "new smart cities." In the case of the former, existing cities are updated with digital infrastructures, including electronic networks, smart devices, and AI-powered applications. In the latter case, entirely new cities are constructed with state-of-the-art digital systems from the outset. These "new smart cities" are often prestigious national projects designed to boast leadership and impress citizens. Examples include Konza City in Kenya, Kitoko Smart City in Kinshasa, and Egypt's New Administrative Capital. These cities are strategically located near existing capitals and are built by coalitions of national and international investors.

The "smarter city" model often builds on grassroots innovation and collaboration between local tech entrepreneurs and municipal governments. Here, AI technologies are integrated into existing urban fabrics, often resulting in vernacular forms of smartness (Pype, 2017) that reflect local social, economic, and political realities. The success of these initiatives depends on their ability to align with local needs and values, rather than imposing a one-size-fits-all model of smart urbanism.

My central argument on urban AI in Africa is that these technologies should be tools for achieving the "good life" rather than ends in themselves. Smartness and connectivity must translate into tangible improvements in people's lives—enhancing mobility, healthcare, education, and economic opportunities.

A Growing Urban Economy

In Africa, urban AI is emerging and has become an important component of the various tech labs around the continent. The consequence is that the economic landscape in urban Africa is rapidly evolving. Startups, incubators, and accelerators have become hubs of innovation and aspiration, supported by a mix of public and private funding.

For instance, in April 2025, the Ghanaian startup Kofa raised $1.8 million to launch an AI-powered battery-swapping network for electric motorcycles. This initiative addresses the heavy reliance on petrol-powered motorcycles and backup generators in Ghana by offering a clean alternative that reduces emissions and lowers fuel costs. Investors include E3 Capital, Injaro, the Shell Foundation, and the UK government, highlighting the global interest in African tech solutions. A whole new language appears in this economic field, e.g., “strategic angels” and “sector veterans,” indicating the particular seniority of investors. The startup will produce AI-powered batteries that can be used by electric motorcycles—one of the most important means of mobility in African cities.

As mentioned in one of the press reports, AI-powered battery-swapping is important in Africa because “Africa’s cities rely heavily on petrol-powered motorcycles and backup generators. Kofa’s swappable battery network offers a clean, reliable alternative—cutting emissions, saving time, and lowering fuel costs for riders and small businesses” (Labari AI, 2025).

The Kofa initiative is only one of the many AI-related initiatives in urban Africa nowadays. By 2030, AI is projected to contribute over $1.2 trillion to Africa’s GDP (Editor, 2025). The South African and Nigerian states, which lead the way, prioritize AI-powered fintech, healthcare diagnostics, and smart agriculture.

Noteworthy examples of AI-driven innovation can be found in the healthcare sector, particularly in the fight against counterfeit drugs. Nigerian pharmacist Adebayo Alonge, for instance, developed an AI-powered platform for drug authentication, which earned him the Hello Tomorrow Global Challenge award in 2019. Similarly, the Ghanaian startup mPedigree leverages mobile technology to verify drug authenticity via SMS, offering a practical and accessible solution to a widespread public health challenge. In agriculture, AI assists weed control and crop management. In Rwanda, for example, startups insert AI-equipped drones to map weed infestations and guide targeted herbicide applications. The Rwandan government also explores AI to provide real-time advice to farmers. Other innovations include ToumAI, a Moroccan startup addressing the challenges of multilingual and multicultural customer service environments. By focusing on underrepresented languages, ToumAI exemplifies how AI speaks to the multilingual contexts of African urban economies.

The economic universe of urban AI is supported by a diverse array of actors. Embassies, UN agencies like UNITAC, and private companies such as Google, Microsoft, and Huawei are investing in African tech ecosystems. The French BPI (Public Investment Bank) finances ToumAI. UNDP and the G7 launched an "AI Hub for Sustainable Development" in 2024. Microsoft plans to invest $80 billion in African data centers and train millions in AI and cybersecurity (Smith, 2025). Google's Africa Startups Accelerator supports black-led AI startups (The Staff, 2023).

Challenges

In our aforementioned Cairo Charter, we signal the importance of inclusive design in AI technologies. We argue for an AI "smartness" that respects local societies, cultures, and needs. Our emphasis on inclusion and ethics remains urgent, particularly given that significant challenges remain, despite the excitement surrounding the economic opportunities of AI development in urban Africa.

A first major concern is the democratization of AI development. African stakeholders emphasize the need for homegrown solutions that meet the financial capacities of African users. The high cost of hardware, relative to income levels, poses a significant barrier. In Kenya, for example, the cost of GPUs (graphic processing units, the chips that often power AI applications) is 31 times higher than in Germany when adjusted for income (Humeau, 2024, p. 25). Additionally, reliance on foreign cloud services (e.g., Amazon Web, Microsoft Azure, or Google Cloud) increases costs and speed, underscoring the need for local data infrastructure.

Digital literacy presents a second critical challenge. In Egypt, for instance, over 25% of its citizens were estimated to be illiterate as of

2017 (Al-Youm, 2018). Such high levels of digitally illiterate citizens significantly constrain the accessibility, adoption, and overall effectiveness of AI technologies.

Data sovereignty, meaning that data generated in Africa should be stored and governed on the continent, is a third pressing concern. Since early 2025, African leaders have been taking serious steps for the local governance of data generated on the continent. For example, the Africa Declaration on Artificial Intelligence, signed in Kigali in April 2025, is a commitment to establishing national AI strategies and governance frameworks in line with the African Union's vision.

Finally, contextualization of AI smartness and innovations remains crucial in order to design AI products that can gain traction among their intended users. In Kinshasa, local tech inventors often frame their innovations as responses to social precarity rather than efficiency (Pype, 2021). Kinshasa's tech developers understand the "problem" of Kinois social life as one of "entrapment." Distrust in others pushes them to invent tech solutions that extract a human intermediary from a financial or information exchange, and replace it with a non-human actor such as an app or a software program. "Tech solutionism" therefore does not carry the same meaning or implications across different global contexts. Furthermore, we need to remain attentive to "other," rather than solely "smart" forms of thinking, moving, and being in the city that already exist when high-tech forms of smartness are introduced into African urban spaces. I argue that a commitment to local urban AI requires a recognition of the fact that technological expertise, including AI knowledge, coexists with other forms of urban knowledge. Such a perspective calls for a more nuanced understanding of how AI is embraced, adapted, or resisted.

In conclusion, urban AI in Africa brings together technology, politics, economics, and culture. An AI hype (Neumark, 2024) is clearly in the air, reflecting the enthusiasm of many African leaders, entrepreneurs, and citizens for AI's potential to improve urban life. Yet, scholarship on urban AI in Africa is only slowly emerging. These are exciting times for AI developers, researchers, and students to shape not only the future of AI in Africa but also the evolving futures of African city life.

References

Africa Declaration on Artificial Intelligence (2025, April 4). *Africa Declaration on Artificial Intelligence*. Kigali, Rwanda: Centre for the Fourth Industrial Revolution Rwanda.

Al-Youm, A.-M. (2018, September 8). *Egypt's illiteracy rates stand at 25.8% in 2017: CAPMAS. Egypt Independent*. https://www.egyptindependent.com/egypts-illiteracy-rates-stand-at-25-8-in-2017-capmas/

Editor (2025, May 19). *The future of AI in Africa: Predictions and visions for 2030. iAfrica*. https://iafrica.com/the-future-of-ai-in-africa-predictions-and-visions-for-2030/

El Khateeb, S., Sherman, S., Pype, K., & Hendawy, M. (n.d.). *The Cairo Charter: A framework for social and environmental justice-driven urban AI in Cairo & beyond* [Research report]. KU Leuven, Faculty of Social Sciences.

Humeau, E. (2024, July). *AI for Africa: Use cases delivering impact. Kenya deep dive* [Report]. GSMA & Axum. https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-for-development/wp-content/uploads/2024/07/KENYA_AIforAfrica.pdf

Labari AI (2025). *Kofa raises $8.1M to scale AI-powered battery-swapping in urban Africa. TL Tech Labari*. https://techlabari.com/kofa-raises-8-1m-to-scale-ai-powered-battery-swapping-in-urban-africa/

Neumark, T. (2024). Hypeful worlds: Putting the brakes on critique in Tanzanian technoscience. *American Ethnologist, 51*(4), 527–539. https://doi.org/10.1111/amet.13144

Pype, K. (2017). Smartness from below: Variations on technology and creativity in contemporary Kinshasa. In C. Mavhunga (Ed.), *What do science, technology, innovation mean from Africa?* (pp. 97–115). MIT Press.

Pype, K. (2021). Digital creativity and urban entrapment in Kinshasa: Experiments in solving precarity. *City & Society, 33*(2), 324–345. https://doi.org/10.1111/ciso.12357

Smith, B. (2025, January 3). *The golden opportunity for American AI*. *Microsoft on the Issues*. https://blogs.microsoft.com/on-the-issues/2025/01/03/the-golden-opportunity-for-american-ai/

The Staff (2023, August 24). *Google launches AI accelerator for Africa*. *Connecting Africa*. https://www.connectingafrica.com/innovation-hub/google-launches-ai-accelerator-for-africa

AI in City Governments – Hype or Mainstream?

Joep Crompvoets

AI has become a ubiquitous presence in our societies, sparking considerable interest and high expectations, particularly regarding its potential applications in the public sector and urban governance. In recent years, city governments have been encouraged to explore AI solutions as part of their digital transformation agendas. However, despite the enthusiasm, there are growing indications that AI's perceived potential in cities may be inflated. As Luusua et al. (2022) note, much of the current discourse around AI carries the hallmarks of a hype, marked by exaggerated claims and disproportionate media attention. In practice, the uptake of AI in city administrations has so far fallen short of these lofty expectations. The gap between the imagined future and present-day implementation raises a critical question: Is AI in urban governance truly a transformative force in the making—a "hit"—or is it still more of a hype? If the latter, we must reflect carefully on the rationale for investing in AI at the city level and identify the key challenges that continue to hinder meaningful adoption.

Defining AI for Urban Governance

This essay explores the current and potential use of AI in city and municipal governments, with a particular focus on the barriers that urban administrations face in adopting these technologies. To frame this discussion, it is first necessary to clarify what is meant by AI in this context. According to the European Union's AI Act (European Parliament & Council of the European Union, 2024), an "artificial intelligence system" refers to software developed using techniques such as machine learning, logic- and knowledge-based approaches, or statistical methods. These systems are designed to achieve human-defined objectives by producing outputs, such as predictions, recommendations, or decisions that influence the environments in which they operate.

In this sense, AI can be considered a key enabling technology for smart cities. It holds promise for enhancing public services, optimizing urban management, and supporting data-driven decision-making for citizens, businesses, and municipal authorities alike. Yet the degree to which AI has become embedded in everyday governance, and the institutional and infrastructural readiness to support it, remains uneven and contested. This essay seeks to critically assess these dynamics and contribute to a more grounded understanding of AI's place in city governments today.

Symbolic vs. Connectionist AI

Two major schools of thought have shaped the development of AI over the past decades: the symbolic and the connectionist approaches (Heratha & Mittal, 2022; Mira, 2008). These paradigms represent fundamentally different visions of how to achieve machine intelligence.

Symbolic AI, often associated with so-called symbolists, seeks to build intelligent systems by explicitly coding logical rules and structured representations of the world. This approach relies on declarative knowledge, such as facts, objects, and axioms, that can be understood and manipulated through reasoning. Its advantages include transparency (the "glass box" effect), explainability, low data requirements, determinism, and human oversight. However, symbolic AI has also been criticized for being rigid, difficult to scale, and reliant on extensive manual modeling of complex real-world domains. It also requires ongoing maintenance and agreement on formalized representations. Common applications include logic-based systems, ontologies, semantic web technologies, rule-based languages, and data modeling tools.

Connectionist AI, in contrast, focuses on learning from data. Rather than preprogramming knowledge, connectionist approaches rooted in neural networks and inspired by biological cognition allow systems to infer patterns and associations through data-driven processing. The strengths of this school lie in its capacity for learning, scalability, adaptability, and ability to handle uncertainty. It is less constrained by human bias at the input level and well-suited to environments characterized by rapid change. Key techniques include machine learning, pattern recognition, backpropagation, genetic algorithms, and deep learning. Nonetheless, connectionist AI has its drawbacks: its models often function as "black boxes" with limited explainability, they require large volumes of data, and they are stochastic, data-biased, and non-deterministic in nature.

With the recent breakthroughs in LLMs such as ChatGPT, public and professional attention has shifted significantly toward connectionist approaches. However, for AI to deliver on its full potential, particularly in complex socio-political contexts like cities, it would be beneficial to explore hybrid models that combine the structured reasoning of symbolic AI with the adaptive capabilities of connectionist systems.

The Public Sector as an AI User

In this evolving landscape, the public sector, and especially city governments, are not merely passive observers. As noted by Guenduez and Mettler (2022), public administrations can serve four strategic roles in the AI ecosystem: as regulators, enablers, leaders, and users. While much of the policy debate has focused on regulation and governance, the role of governments as users of AI remains underexplored, despite its enormous potential to enhance city policymaking, service delivery, and internal operations (van Noordt & Misuraca, 2022; Manzoni et al., 2022). AI can support various phases of the policy cycle, including early detection of social issues, modeling potential impacts of policy options, expediting decision-making processes, monitoring implementation, and evaluating outcomes. In terms of public service delivery, AI systems can improve information access, personalize services, automate routine interactions, reduce administrative burdens, generate new types of services, mitigate corruption, and build citizen trust. At the same time, internal operations within city administrations can benefit from AI-driven improvements in human resources management, recruitment, financial oversight, cybersecurity, procurement processes, and fraud detection. Taken together, these opportunities suggest that AI, when implemented responsibly and strategically, can support not only smarter cities but also more responsive, transparent, and efficient local governance. However, realizing this potential will require cities to critically assess both technological options and institutional readiness, while also addressing the ethical, legal, and social implications of AI use in the public domain.

The Implementation Gap

Despite the considerable potential of AI for improving urban governance, the current level of adoption in city governments remains limited. As Ben Rjab et al. (2023) point out, we still know relatively little about how AI is actually being used in municipal contexts. Most existing applications are still in the conceptual or developmental stages, with many of their promised benefits yet to be empirically demonstrated (van Noordt & Misuraca, 2022). With the exception of some basic chatbot implementations, AI adoption in cities is generally still in its infancy.

This underscores a broader truth: the adoption of innovative technologies like AI in the public sector is far from straightforward. Governments and city administrations face a range of challenges, including technological, organizational, legal, ethical, societal, and institutional barriers as well as issues related to trust, awareness, and public procurement. The mere existence of mature AI solutions on the market does not guarantee that public institutions will adopt or integrate them easily, or at all.

Making AI Work for Cities

Nonetheless, the limited but emerging examples of AI use in city governance offer a glimpse into the possible futures of smart cities. These early cases point to ways in which AI might eventually become mainstream, enhancing policy design, public service delivery, and internal management processes. However, to reach that point, substantial efforts are needed to better integrate AI into the real-world workflows and structures of public administrations. Such efforts include resolving technical and ethical concerns, bridging the strengths of both symbolic and connectionist AI approaches, building individual and institutional capacities, and sharing concrete examples of successful implementation.

These actions can help foster trust among civil servants, policymakers, citizens, and businesses, shifting perceptions of AI from a source of risk to a credible opportunity for improvement and innovation. In sum, the current state of AI in city governments suggests that the initial hype phase is receding. While AI is not yet a mature, mainstream technology in urban governance, it is gradually positioning itself to become one. Realizing this potential will require moving beyond inflated expectations and approaching AI not as a silver bullet, but as a complex tool whose value depends on how thoughtfully and responsibly it is deployed.

References

Ben Rjab, A., Mellouli, S., & Corbett, J. (2023). Barriers to artificial intelligence adoption in smart cities: A systematic literature review and research agenda. *Government Information Quarterly, 40*(1), 101814. https://doi.org/10.1016/j.giq.2023.101814

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

Goel, A. K. (2021). Looking back, looking ahead: Symbolic versus connectionist AI. *AI Magazine, 42*(1), 83–85. https://doi.org/10.1609/aaai.12026

Guenduez, A. A., & Mettler, T. (2022). Strategically constructed narratives on artificial intelligence: What stories are told in governmental artificial intelligence policies? *Government Information Quarterly, 39*(3), 101719. https://doi.org/10.1016/j.giq.2022.101719

Heratha, K., & Mittal, M. (2022). Adoption of artificial intelligence in smart cities: A comprehensive review. *International Journal of Information Management Data Insights, 2*, 100076. https://doi.org/10.1016/j.jjimei.2022.100076

Luusua, A., Ylipulli, J., Foth, M., & Aurigi, A. (2022). Urban AI: understanding the emerging role of artificial intelligence in smart cities. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01537-5

Manzoni, M., Medaglia, R., Tangi, L., van Noordt, C., Vaccari, L., & Gattwinkel, D. (2022). *AI Watch. Road to the adoption of Artificial Intelligence by the public sector* (JRC129100). Publications Office of the European Union. https://doi.org/10.2760/288757

Mira, J. (2008). Symbols versus connections: 50 years of artificial intelligence. *Neurocomputing, 71*(13–15), 671–680. https://doi.org/10.1016/j.neucom.2007.06.009

van Noordt, C., & Misuraca, G. (2022). Artificial intelligence for the public sector: Results of landscaping the use of AI in government across the European Union. *Government Information Quarterly, 39*(3), 101714. https://doi.org/10.1016/j.giq.2022.101714

The Smart Library as a Microcosm of China's Smart City Vision

Chao Wang

Introduction

In recent years, the Chinese government (The State Council of the People's Republic of China, 2023; 2024) has strongly supported the development of smart cities through national strategies such as the "Digital China" initiative and the "New Urbanisation Plan". These policies aim to apply digital technologies across public services, administrative systems, and city management. While much of the existing research has focused on prominent areas like transportation, energy, and infrastructure, less attention has been given to more local and cultural aspects of the smart city landscape.

Among these less studied areas, the smart library is a particularly important but often overlooked example. Although it belongs to the cultural and educational sector, the smart library reflects the broader principles of digital governance. It not only provides access to knowledge and supports cultural participation, but also uses digital technologies to manage public spaces and influence how people interact with information (Leorke et al., 2018).

This essay argues that smart libraries in China are more than just service hubs or symbols of technological progress. They reflect the broader vision of the smart city. By looking at smart libraries, we see how digital tools shape public life, manage data, and influence citizen behaviour in line with the state's goals for modernisation. Using the concept of socio-technical imaginaries—shared ideas of the future shaped by technology (Mager & Katzenbach, 2021)—this essay shows how smart libraries serve both practical and symbolic roles in China's push for digital governance.

Policy Context: Smart Cities with Chinese Characteristics

China's smart city strategy is part of a broader national push to modernise governance through digital technology. Policies like the 14th Five-Year Plan for the Digital Economy outline goals for building a digitally empowered society (NDRC, 2021). In practice, the strategy blends top-down planning with cooperation between the state, tech firms, universities, and research institutes. Efforts focus on creating digital platforms, simplifying administration, and improving public services in urban areas. Within China's digital policy framework, smart libraries are seen as important parts of both "smart governance" and "smart culture".

Once mainly focused on education and culture, libraries now also promote digital literacy and support access to online government services (Luterek, 2020). This positions them within the broader system of digital governance, where they help manage information and reinforce public norms. Their dual role—as cultural centres and tools for information control—shows how public spaces are being reshaped by the smart city agenda (Cao et al., 2018). Far from being a minor innovation, the smart library reflects how technology is used to align cultural institutions with national goals of efficiency, control, and integration.

Smart Libraries: Civic Interface or Technological Showcase?

Smart libraries in China increasingly adopt technologies such as self-service machines, AI chatbots, AR/VR environments, and facial recognition systems to enhance service efficiency and user experience (Li & Yuan, 2021; Li, 2024). These developments reflect broader efforts to integrate libraries into smart city infrastructure, framing them as modern and tech-enabled public spaces. Beyond convenience, smart libraries also serve as entry points through which citizens encounter digital tools and norms, promoting a "smart lifestyle" and supporting adaptation to data-driven governance (Dong et al., 2022; Liu et al., 2024).

A key civic role of smart libraries is to promote digital literacy, particularly among urban adults and seniors. While training programmes help users access e-government services or navigate mobile payments, they often emphasise functional use over critical engagement with issues like data privacy or algorithmic bias (Elgamal et al., 2024). Furthermore, access to these services remains uneven, and vulnerable groups—such as older or less tech-savvy individuals—may be unintentionally excluded (Wang et al., 2023).

Case studies from Shenzhen, Hangzhou, and Tianjin illustrate varied approaches: some libraries emphasise interactivity and youth engagement, while others focus on automation or civic education (Zhao, 2021; Liu, 2023). Despite these differences, a common challenge remains: balancing technological display with meaningful participation. This raises a broader question: do smart libraries truly foster civic engagement, or are they primarily showcases for digital innovation? Answering this is essential to understanding their evolving role in China's smart city development.

Many smart library programmes in China focus on practical digital skills, like using smart devices, but often overlook critical digital awareness. In contrast, the EU's DigComp and cross-country research such as the ySKILLS and REMEDIS research projects highlight media literacy, data literacy, and critical thinking about algorithms (Spurava & Kotilainen, 2022; Van Audenhove et al., 2024). From this perspective, digital literacy means not just adapting to technology but also being able to question and assess it. This broader approach highlights the potential of libraries to support more informed and empowered digital citizens.

Critical Reflections: Surveillance, Digital Literacy, and Participation

Although smart libraries in China are praised for their efficiency and innovation, they also reveal tensions between technology, governance, and public life. Surveillance tools like facial recognition, behaviour tracking, and usage analytics are often used in the name of service improvement. While these tools may boost efficiency, they raise serious concerns about privacy and consent, especially as many users are unaware of how their data is collected or used (Chen, 2023; Xu & Shang, 2024). This reflects a wider trend in China's smart governance, which depends on centralised data collection (Yu, 2024).

Digital literacy presents another challenge. Training programmes in both schools and libraries often focus on basic technical skills and risk prevention, offering little space for critical reflection on how digital systems function or monetise user data (Chai, 2025; Zheng, 2025). While libraries help users adopt new technologies, these efforts typically emphasise system compliance over critical understanding (Ince, 2022). Moreover, access remains uneven, with older adults and less digitally skilled individuals frequently excluded, thus risking a reinforcement of the digital divide (Wang et al., 2023).

The notion of "public participation" in smart libraries also deserves critical attention. While users may interact with digital services, these engagements are often limited and highly structured. Genuine opportunities for input or shared decision-making are rare, making participation largely symbolic and geared towards digital conformity rather than empowerment (Buyannemekh et al., 2024). These concerns—surveillance, exclusion, and limited agency—complicate the optimistic view of smart libraries. A more balanced approach is needed, one that prioritises transparency, inclusion, and ethical responsibility alongside technological progress.

China's model of smart cities relies on centralised data collection and top-down integration, while European approaches tend to be more decentralised and citizen-focused. Initiatives under Horizon Europe and Urban Innovation Actions (UIA), for example, support smart infrastructure that emphasises digital inclusion and civic co-design (Campillo-Alhama et al., 2024). The Helsinki Oodi Library illustrates this approach, offering a participatory public space with minimal surveillance and strong adherence to GDPR standards (Mady & Hewidy, 2025). In contrast, China's model, despite its technological sophistication, raises concerns about user agency and data transparency. These differences point to deeper divergences in civic values and ideas of public space. Without civic grounding, algorithmic governance risks weakening the democratic role of institutions like libraries.

Conclusion: Rethinking the Smart City Through the Library Lens

Smart libraries offer a concrete example of how China's smart city vision is put into practice in everyday public spaces. By blending digital technology with cultural services and data management, they bring smart city principles to the local level. However, their success depends on inclusivity. Current models often favour digitally skilled users, leaving behind those with limited access or digital literacy. Moreover, extensive data collection—if poorly regulated or explained—can erode public trust and compromise the library's civic role. To remain inclusive, smart libraries must ensure transparency, address algorithmic bias, and foster meaningful public participation.

China's experience with smart libraries highlights both the promise and the challenges of digital innovation in public spaces. These libraries illustrate how digital governance is reshaping everyday civic life, offering new services while raising concerns about equity, privacy, and participation. Viewed through the lens of the library, the smart city becomes more than a technological project; it is a site where competing values are negotiated. While China's model prioritises centralised control and state-led infrastructure, European approaches tend to emphasise civic co-creation, data rights, and institutional trust. Rather than framing these models as opposites, future research and policy should encourage dialogue between them, seeking a balance between efficiency and inclusion, innovation and accountability. In this light, smart libraries emerge not only as service institutions but also as key arenas for shaping the civic meaning of digital governance.

References

Buyannemekh, B., Gasco-Hernandez, M., & Gil-Garcia, J. R. (2024). Fostering smart citizens: The role of public libraries in smart city development. *Sustainability, 16*(5), Article 1750. https://doi.org/10.3390/su16051750

Campillo-Alhama, C., Santa-Soriano, A., & Torres-Valdés, R. M. (2024). Citizen science and artificial intelligence in Horizon 2020 and Horizon Europe projects: Communication and scientific impact. *Profesional de la Información, 33*(4), Article e330417. https://doi.org/10.3145/epi.2024.0417

Cao, G. H., Liang, M. L., & Li, X. G. (2018). How to make the library smart? The conceptualization of the smart library. *The Electronic Library, 36*(5), 811–825. https://doi.org/10.1108/EL-11-2017-0248.

Chai, L. (2025). Media literacy in China: The impact of education on individuals' media behavior. In *Proceedings of the 3rd International Conference on Social Psychology and Humanity Studies*. https://doi.org/10.54254/2753-7048/85/2025.21300

Chen, J. N. (2023). Research on the reading promotion model of universities based on smart libraries. *International Journal of New Developments in Education, 5*(11), 59–63. https://doi.org/10.25236/IJNDE.2023.051112

Dong, L., Zhu, F., Meng, Y., & Guan, R. (2022). Cultivation strategies for users of smart library in China. In *Proceedings of the 2022 International Workshop on Control Sciences and Automation (WCSA)* (pp. 48–52). ACM. https://doi.org/10.1145/3543303.3543319

Elgamal, R., La Rose, T., Detlor, B., Julien, H., & Serenko, A. (2024). A community partnership approach to digital literacy training for older adults between public libraries and seniors' organizations. *Canadian Journal of Information and Library Science, 47*(1), 3–17. https://doi.org/10.5206/cjils-rcsib.v47i1.16593.

Ince, F. (2022). Digital literacy training: Opportunities and challenges. In M. Tuncer (Ed.), *Handbook of research on the role of libraries, archives, and museums in achieving civic engagement and social justice in smart cities* (Chapter 9). IGI Global. https://doi.org/10.4018/978-1-7998-8363-0.ch009

Leorke, D., Wyatt, D., & McQuire, S. (2018). "More than just a library": Public libraries in the 'smart city'. *City, Culture and Society, 15*, 37–44. https://doi.org/10.1016/j.ccs.2018.05.002.

Li, D. Y. (2024). Adoption of artificial intelligence in public and private libraries of China: Determinants, challenges, and perceived benefits. *Profesional de la Información, 33*(4), e330416. https://doi.org/10.3145/epi.2024.ene.0416.

Li, M. M., & Yuan, Q. (2021). Construction of VR service platform of smart library. In *Proceedings of the 2021 4th International Conference on Information Systems and Computer Aided Education (ICISCAE)* (pp. 2313–2317). ACM. https://doi.org/10.1145/3482632.3484153

Liu, L. L. (2023). Strategy research of building smart library. *Academic Journal of Humanities & Social Sciences, 6*(21), 84–88. https://doi.org/10.25236/AJHSS.2023.062114

Liu, Y., Huang, J. J., & Wang, Y. H. (2024). Digital transformation and the future of public libraries in China: Assessing innovations, challenges, and user engagement. *Studies in Social Science & Humanities, 3*(5), 72–79. https://www.paradigmpress.org/SSSH/article/view/1146

Luterek, M. (2020). Public libraries: Their role in smart city strategies. *Zagadnienia Informacji Naukowej, 58*(2), 27–44. https://doi.org/10.36702/zin.732

Mady, C., & Hewidy, H. (2025). The public library building as nexus for social interactions: Cases from Helsinki. *City, Culture and Society, 40*, Article 100610. https://doi.org/10.1016/j.ccs.2024.100610

Mager, A., & Katzenbach, C. (2021). Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified. *New Media & Society, 23*(2), 223–236. https://doi.org/10.1177/1461444820929321

National Development and Reform Commission (NDRC) of the People's Republic of China. (2021). *Resolution on the outline of the 14th five-year plan for economic and social development and long-range objectives through the year 2035*. https://en.ndrc.gov.cn/policies/202203/P020220304367370277643.pdf

Spurava, G., & Kotilainen, S. (2022). Young people as empirical experts of participatory research in the age of information disorder. *Seminar.net: Media, Technology & Lifelong Learning, 18*(1), 1–15. https://doi.org/10.7577/seminar.4740

The State Council of the People's Republic of China. (2023). *China unveils plan to promote digital development*. The State Council, The People's Republic of China. https://english.www.gov.cn/policies/latestreleases/202302/28/content_WS63fd33a6c6dd0a757729e752c.html

The State Council of the People's Republic of China. (2024). *China releases five-year action plan for people-centered new urbanization*. The State Council, The People's Republic of China. https://english.www.gov.cn/policies/latestreleases/202407/31/content_WS66aa26c8c6dd0868f4e8e99f2.html.

Van Audenhove, L., Vermeire, L., Van den Broeck, W., & Demeulenaere, A. (2024). Data literacy in the new EU DigComp 2.2 framework: How DigComp defines competences on artificial intelligence, internet of things and data. *Information and Learning Sciences, 125*(5/6), 406–436. https://doi.org/10.1108/ILS-06-2023-0072.

Wang, W. H., Zhang, Y. T., & Zhao, J. J. (2023). Technological or social? Influencing factors and mechanisms of the psychological digital divide in rural Chinese elderly. *Technology in Society, 74*, 102307. https://doi.org/10.1016/j.techsoc.2023.102307.

Xu, X. B., & Shang, J. (2024). Research on the construction scheme of smart library based on blockchain technology. *Measurement: Sensors, 31*, Article 100943. https://doi.org/10.1016/j.measen.2023.100943

Yu, M. L. (2024). Key challenges and potential solutions in smart city governance. *Scientific and Social Research, 6*(8), 173–180. https://doi.org/10.26689/ssr.v6i8.7818

Zhao, L. M. (2021). Construction of library's smart ecosystem in the "Smart+" era. *International Journal of Frontiers in Sociology, 3*(1), 52–59. https://doi.org/10.25236/IJFS.2021.030107.

Zheng, S. S. (2025). A study on digital literacy of Chinese secondary school teachers in the era of artificial intelligence. *International Journal of Education and Social Development, 2*(1), 48–51. https://doi.org/10.54097/tdy31h37.

Debate Summary

AI in the City: Smart Governance, Civic Agency, and Ethical Futures

Leen d'Haenens, Florence Chee, Kurt Feyaerts, Katrien Pype, Chao Wang, Vangelis Palaskas, & Bruno Dupont

The rapid integration of AI into urban governance has sparked both enthusiasm and critical reflection worldwide. Whether used to monitor noise levels in Leuven, model climate adaptation through digital twins, or manage infrastructure in Kinshasa and Cairo, AI is shaping urban life in diverse and often uneven ways. These developments reveal the complex balancing act that cities face: between the drive for efficiency and the need to support community-led innovation, between decisions shaped by technical expertise and those shaped through public participation, and between global ethical guidelines and the lived moral values of local communities. This debate summary reflects on these challenges through a range of international case studies viewed through an interdisciplinary lens. From smart libraries in East Asia to surveillance systems in African megacities, each example shows how AI in cities is shaped and determined by unique political goals, technological resources, and cultural contexts. Instead of one model of a smart city, these cases point to multiple, often conflicting, pathways that are unfolding around the world.

A noteworthy example is the city of Leuven, which is integrating AI and digital tools to improve public health, urban mobility, and climate resilience. Policy advisor Tim Guily explained that AI is viewed not as an end in and of itself. Instead, it is seen as a part of a broader toolbox, one used to tackle complex urban challenges. Projects include using sensors and real-time data for traffic monitoring, deploying digital twins for environmental planning, and encouraging citizen involvement to ensure locally relevant solutions. Interestingly, across contexts, the biggest challenges are not technical but deeply political: how to embed AI in ways that are democratic, inclusive, and ethically grounded. This essay concludes with recommendations for policymakers and practitioners committed to ensuring that urban AI strengthens civic agency and public trust, instead of undermining it.

Context Matters: AI Is Always Local

Urban AI is not a monolith. The "smartening" of Kinshasa, for instance, unfolds under radically different infrastructural and political conditions than Seoul or Leuven. Katrien Pype (in this book) reminds us that informal innovation plays a central role in many African cities: fablabs, bottom-up tech experiments, and hybrid governance ecosystems. These grassroots spaces often reflect a form of digital creativity that is adaptive and deeply local. Yet, they are seldom recognized or integrated into formal policy structures. This disconnect raises the question of how AI systems might be built to support, rather than displace, such local practices.

In the city of Leuven, a highly institutionalized setting, AI is embedded into data models that assess air quality, road conditions, and crowd dynamics. The city's roadmap for smart innovation prioritizes digital twins, crowdsourced monitoring, and actionable feedback loops. Yet even here, gaps remain. Despite the apparent robustness of these systems, citizen involvement in defining the problems, interpreting the data, and shaping the outcomes is not always guaranteed. The question of who gets to define urban priorities lingers across both contexts.

These contrasting examples provide a fundamental insight: AI must be understood and implemented through local lenses. Exporting smart city templates risks flattening the complex textures of urban life. The digital infrastructure of a city in the Global North may not, and often should not, be replicated in the Global South without fundamental rethinking. Local histories, socio-economic structures, and governance capacities must all inform the design of AI systems.

Beyond Technocracy: Ethical Frameworks for AI

Florence Chee brings ethical nuance into focus, cautioning against the technocratic logic that often underpins AI deployments, where efficiency, predictability, and control tend to dominate (Burris, 1989; Zysman & Nitzberg, 2020). She argues that ethical deliberation should not be reduced to regulatory checklists or abstract principles. Instead, ethics must be grounded in lived experiences and shaped by the social and cultural values of specific communities. Several contributors emphasized the need for plural ethical design (Chee, 2018). While global institutions may provide overarching articulations of rights, dignity, and care, these must be complemented by locally rooted moral vocabularies. In urban contexts, collective notions of responsibility, care, and reciprocity can provide a much-needed counterbalance to individualistic or market-driven approaches. The ethical stakes are particularly high in urban governance, where AI increasingly influences the distribution of resources and opportunities (Pulijala, 2024).

Should a noise-detection algorithm be used to suppress protest or to support vulnerable residents? Should traffic optimization prioritize commercial logistics or pedestrian well-being? These are normative ones that demand inclusive and context-sensitive deliberation about what exactly constitutes a just and livable city.

Data, Power, and the Urban Condition

The conversation around data ownership and use has sparked sharp critiques of what some scholars term *data colonialism*, a concept that draws attention to the ways in which data extraction mirrors earlier forms of imperial domination (Gray, 2023). In the context of urban AI, this critique becomes especially urgent. Sensors, facial recognition systems, and behavioral tracking technologies often collect personal data under conditions that lack transparency or meaningful consent. As these systems become embedded in the routines of city life, they risk reinforcing power asymmetries between those who generate data and those who govern it (Farooq et al., 2023).

Examples from cities across the globe underscore this concern. In China, smart libraries reflect the wider goals of digital governance. They use technologies such as facial recognition and data analytics to improve services, but also raise important questions around privacy, literacy, and public participation (see Chao Wang in this book). Despite their technological sophistication, such systems often privilege surveillance and automation over dialogue and public accountability. Similar patterns emerge elsewhere: in Kinshasa and Ghent, and even in Leuven, where telecom data is used to evaluate the economic impact of cultural events by tracking visitor flows and spending patterns. While such practices may support evidence-based policymaking, they risk framing residents primarily as economic units instead of active, rights-bearing citizens.

The implications go far beyond individual privacy. At stake are broader questions about who controls urban data infrastructures, who benefits from them, and who is left exposed or excluded. Scholars have shown how algorithmic systems can reproduce existing racial, gendered, and economic inequalities—especially when they operate without critical, and human, oversight (Browne et al., 2023; Zajko, 2022).

If AI is to contribute meaningfully to democratic urban futures, it must be built on principles of transparency, equity, and shared governance. This means rethinking not just how data is collected, but also how power is distributed through the very systems that claim to optimize our cities.

Participatory Urbanism: Co-Design and Inclusion

Participation is and should be a foundational concern in the development of urban AI. Yet in practice, citizen engagement is often limited to surface-level consultations or digital feedback tools that offer little influence over decision-making. For urban AI to be truly democratic, participation must go deeper: it must involve communities in both identifying challenges and shaping solutions, as well as assessing its outcomes. A recurring theme in the urban AI conversation was the need for co-design processes that are inclusive and context-sensitive. Such approaches recognize citizens as active partners in creating and adopting the technologies that (will) affect their daily lives. This includes creating what have been called "spaces of consent," settings where communities can openly deliberate the purpose, boundaries, and implications of technological systems (Chee, 2022). These participatory spaces are especially important where trust in public institutions is already strained. In Kinshasa, for example, the rollout of digital voting machines sparked skepticism, which is more a reflection of broader concerns about state legitimacy and accountability than of the technology itself. Even the most sophisticated AI systems are unlikely to succeed if they are introduced without public trust or meaningful dialogue.

Inclusive design also requires attention to those often left out of digital governance: older adults, people with limited digital access or literacy, activists, and marginalized communities. It calls for providing analog alternatives to digital tools, designing for accessibility, and embedding principles of equity at every stage of technological development. Participation, in this sense, is an ongoing, evolving process, one that demands commitment, creativity, and humility from designers, policymakers, and technologists alike.

Living Cities, Living Data: Heritage and Justice

Cities are complex social environments, grounded in memory, culture, and emotional connection. It goes without saying that they are far more than data networks and cannot be understood through technology alone. Several contributors to the urban AI conversation stressed the importance of recognizing and preserving "living heritage:" the murals, protest signs, soundscapes, informal gatherings, and fleeting moments that embody the spirit of a city. Within this context, *visual justice* emerges as a critical lens (Mayrhofer-Hufnagl & Ennemoser, 2023). We considered the questions: Who is visible in the data? Whose stories are being told? Which experiences are acknowledged, and which are erased? In Melbourne, activists have used digital tools to map protest art as a form of resistance against the silencing of dissent. For example, in Ghent, researchers employ sound modeling to explore how urban modifications impact the auditory experience in public spaces, raising questions about which senses and sensitivities are prioritized.

This line of thinking, where urban data is treated as a form of heritage, invites a fundamental shift: from control to care. As such, smart cities can be envisioned as places that both honor shared histories and support everyday storytelling, to nurture belonging among their residents. This shift calls for new kinds of metrics, new modes of engagement, and new alliances between technologists, artists, anthropologists, planners, and communities, reminding us that a truly intelligent city is not only efficient, but also attentive to the histories and hopes of those who inhabit it.

In the field of living heritage, a specific case (challenge?) discussed among several participants is the systematic documentation of a city's linguistic landscape, which comprises the visible (non-)verbal signs in public spaces as apparent on signs, billboards, graffiti, murals, storefronts, public transport, stickers, etc. and which reflects various aspects of a city's or neighborhood's cultural diversity, migration patterns, language policies, and social dynamics (Gorter & Cenoz 2024). Integrated into urban space, linguistic landscapes represent an omnipresent and ever-changing aspect of living sociocultural heritage, expressing the voices of a city's past and present (groups of) inhabitants, students, workers, and visitors. In light of the three major challenges to systematically document a city's linguistic landscape (space: *how to systematically map all signs of an entire city?*, time: *how to keep track of constant changes?*, and management: *how to collect, annotate, analyze, and disclose the huge amount of LL-signs?*), this ambitious endeavor represents a formidable challenge and opportunity to integrate AI-driven technology (for instance, smart glasses equipped with state-of-the-art recognition and localization software, Gilles & Ziegler 2021) with various aspects of citizen engagement. Despite the huge capabilities and potential of urban AI and computational technologies, crowdsourcing through the active participation of (groups of) citizens is required in order to achieve both spatial and temporal sustainability. Many different levels and dimensions of a city's social fabric such as neighborhood or street committees, shop owners and shopping streets, families and tourists, graffiti taggers, but definitely also schools, retirement homes, or even a city's postal or garbage service may be motivated to play a crucial role in the realization of this massive endeavor. Here as well, context matters, meaning that AI requires complementary crowdsourcing through different types of social structures and organizations in cities in different parts of the world. Regardless of the specific modalities of its regional or sociocultural implementation, this type of inclusion project appeals to what previously was referred to in terms of a "foundational concern," requiring citizens to actively participate at multiple stages of the project: not only to provide input data, but also to participate in the local semantic enrichment and regular updating of the data collection.

Also, in terms of data ownership and use, the citizens' multi-faceted involvement, alongside urban services and AI, in the documentation of a city's linguistic landscape provides the strongest possible prevention of data colonialism. In light of these observations, the envisaged systematic mapping of urban linguistic landscapes provides a challenge par excellence to balance and integrate a smart city's ambition with respect to the implementation of urban AI with concerns about genuine civic agency and added ethical value.

Reflections and Recommendations

Urban AI is a civic and ethical challenge. As such, we need to reframe the question from what AI *can* do to what it *should* do. Meeting that challenge requires a multi-dimensional and grounded approach.

First, *AI strategies must be context-specific*. The differences between Kinshasa, Seoul, Ghent, and Leuven are, apart from infrastructural, above all cultural, political, and epistemological. One-size-fits-all solutions risk misalignment with local needs and values. Urban AI initiatives should develop from local contexts and be designed through ongoing dialogue with the communities they aim to support.

Second, *ethics must be embedded from the outset*. On a basic level, this means adding ethicists to existing projects. We need to take it a step further and engage deeply with the cultural values, social norms, and philosophical traditions that shape our understanding of justice, dignity, and responsibility. Ethical reflection must therefore inform design choices rather than follow them.

Third, *data ownership and governance must move to the center of urban AI policymaking*. Citizens have the right to know how their data is collected, processed, shared, and stored. Stronger transparency frameworks are needed, along with support for alternative models such as data cooperatives or citizen-led trusts that give communities greater control over their digital environments.

Fourth, *participation must be institutionalized*. Cities should establish permanent, well-resourced forums for citizen deliberation on technological futures; spaces that integrate both digital and analog engagement, ensuring that diverse voices are empowered to take part in shaping outcomes.

Fifth, *urban AI must attend to the sensory, cultural, and emotional textures of city life*. Initiatives that document protest art, ambient soundscapes, or shifting public rituals should be recognized as vital forms of civic intelligence. Balancing innovation with cultural memory and aesthetic justice enriches our understanding of what cities are, and what they can become.

Finally, *governments and public institutions must build internal capacity*. Civil servants need training in both AI deployment and its broader social and ethical implications. Procurement practices should favor public-interest technologies, and evaluation frameworks must expand beyond efficiency and return on investment to include inclusion and democratic accountability.

In conclusion, the challenge lies not merely in building smart cities that serve elite convenience, but in shaping urban innovation to support collective flourishing, guided by human dignity, ecological sustainability, and democratic participation. The future of urban AI will not be determined solely by algorithms but by the choices communities make together about the kind of cities they wish to inhabit.

References

Browne, J., Cave, S., Drage, E., & McInerney, K. (2023). *Feminist AI*. Oxford University Press.

Burris, B. H. (1989). Technocratic organization and control. *Organization Studies*, *10*(1), 1–22. https://doi.org/10.1177/017084068901000101

Chee, F. (2018). An Uber ethical dilemma: examining the social issues at stake. *Journal of Information, Communication and Ethics in Society, 16*(3), 261–274. https://doi.org/10.1108/JICES-04-2018-0042

Chee, F. (2022). Communication as conscience. *communication +1, 9*(2), Article 7. https://doi.org/10.7275/8v86-2k35

Farooq, S., Farooq, B., Basheer, S., & Walia, S. (2023). Balancing environmental sustainability and privacy ethical dilemmas in AI-enabled smart cities. In *Advances in environmental engineering and green technologies book series* (pp. 263–286). https://doi.org/10.4018/979-8-3693-0892-9.ch013.

Gilles, P. & Ziegler, E. (2021). Exploring corpus linguistics approaches in linguistic landscape research with automatic text recognition software. In *Linguistic Landscapes im deutschsprachigen Kontext*. (pp. 65–86). https://doi.org/10.3726/b17795.

Gorter, D. & Cenoz, J. (2024). *A panorama of linguistic landscape studies*. De Gruyter. https://doi.org/10.2307/jj.6947033

Gray, C. (2023). More than extraction: Rethinking data's colonial political economy. *International Political Sociology*, *17*(2). https://doi.org/10.1093/ips/olad007

Mayrhofer-Hufnagl, I., & Ennemoser, B. (2023). Advancing justice in a city's complex systems using designs enabled by space. *International Journal of Architectural Computing, 21*(2), 280–296. https://doi.org/10.1177/14780771231168223

Pulijala, S. (2024). Artificial intelligence in governance: Opportunities, challenges, and ethical implications for public administration. *International Journal For Multidisciplinary Research*, *6*(6). https://doi.org/10.36948/ijfmr.2024.v06i06.29990

Zajko, M. (2022). Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociology Compass*, *16*(3). https://doi.org/10.1111/soc4.12962

Zysman, J., & Nitzberg, M. (2020). Governing AI: Understanding the limits, possibility, and risks of AI in an era of intelligent tools and systems. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3681088

Conclusion

Leen d'Haenens, Ans De Nolf, & Bieke Zaman

As we celebrate 600 years of academic legacy at KU Leuven, we look to the future of our society that is being reshaped to the rhythm of AI developments, considering both the challenges and opportunities it may present. Drawing on diverse disciplinary and regional perspectives, and informed by two polyvocal AI dialogue events, one in New York and one in Leuven, this collection of essays serves as both a testament to the transformative potential of AI and a critical engagement with the ethical, societal, ecological, and political questions surrounding its development, use, and government.

Our book is structured around three central themes: *AI and human employment*, *AI and sustainability*, and *AI and urban life*. It explores the remarkable promise and complex challenges of the rapid advancement of AI technologies from complementary perspectives, drawing on insights rooted in both academic research and the lived experiences of stakeholders and practitioners in the field.

At its best, technological developments that rely on AI have the potential to catalyze innovation, enhance human capabilities, and bring solutions that address urgent societal issues. In areas such as healthcare and education, these AI-driven technologies can enable the provision of personalized instruction through tools that support, rather than replace, human educators. In care work, the use of AI technologies can support routine tasks such as scheduling and documentation, allowing caregivers to devote more time to empathy and personal connection. Across various creative disciplines, GenAI is challenging traditional notions of authorship and the artistic process. In art education, for example, it encourages new forms of reflection and collaboration, enabling students to experiment, iterate, and refine their creative intentions by manipulating prompts and visual references. Rather than undermining creativity, we can use AI in conversations as our generative partner, pushing boundaries, broadening access, and bringing fresh ideas.

Smart cities are another area in which the potential of AI is particularly evident. From predictive policymaking to automated public services, AI is often seen as integral to responsive and efficient urban governance. In China, for example, smart libraries equipped with AI assistants and immersive technologies are seen as a forward-looking and innovative means of engaging citizens with knowledge and digital culture. Meanwhile, new conceptions of African cities are emerging, centered on the idea of vibrant hubs of AI innovation. Localized applications include AI-powered battery-swapping networks for electric motorcycles, drug authentication platforms, precision agriculture tools using AI-equipped drones, and multilingual customer service systems that respond to the continent's unique cultural and linguistic diversity. These technologies are not just functional; they are also aspirational, signaling how AI can be used to address real needs, promote inclusion, and stimulate economic growth.

However, the narrative and imagination of AI as an emancipatory force remain incomplete if we do not account for the ways in which it is shaped by human and societal dynamics, or if we overlook its potentially harmful aspects. The deployment of AI often exacerbates existing inequalities or introduces new forms of exclusion and control. In the workplace, for example, AI can intensify managerial oversight, thereby eroding worker autonomy and undermining job quality. The use of opaque algorithmic systems often prioritizes efficiency over human judgment, sociality, and empathy, while reshaping labor through task redistribution and surveillance. This shift endangers workers in precarious or gig-based employment in particular, as they face heightened instability and diminished rights. Public service infrastructure is also under strain, as digital-only systems often overlook the human and emotional aspects of citizen engagement.

Moreover, we need to reject the illusion of inevitability in technological development, remaining vigilant against dominant narratives that equate innovation with progress without democratic input. To ensure this democratic input, there is a need for more participatory and context-sensitive approaches.

Environmental costs are another cause for concern. The AI industry is resource-intensive and dependent on the extraction of critical minerals, which is often carried out under exploitative conditions in the Global South. The environmental footprint of AI development, training, and use extends far beyond carbon emissions, encompassing energy-intensive data infrastructures, water consumption, and exploitative mining operations that disproportionately impact the Global South. Despite corporate claims of sustainability, the sector frequently engages in greenwashing by promoting techno-utopian narratives that obscure the ecological reality of AI infrastructure. Current regulations are weak, and voluntary corporate standards are insufficient for addressing these systemic harms. To address this, for example, Benedetta Brevini's eco-political economy framework calls for a deeper reckoning with the full lifecycle of AI: from resource extraction to digital waste, arguing that any responsible future for AI must center ecological justice.

AI's ethical complexities extend to the creative fields. Generative models trained on large amounts of online content raise serious questions about intellectual property, artistic integrity, and authorship. These systems' "black box" nature masks the layers of invisible labor (such as that of coders, researchers, and artists) whose contributions are often uncredited and uncompensated. AI-generated outputs are synthetic recombinations of existing human knowledge, far from being wholly original. As such, they blur the boundaries between creativity and computation in ways that require critical scrutiny.

In smart urban environments, the expansion of AI often comes at the expense of privacy and civic trust. Surveillance technologies, from facial recognition to behavioral tracking, are increasingly embedded in public infrastructure, sometimes without citizens' awareness or consent. In contexts such as China's smart libraries, digital literacy training focuses on technical compliance rather than cultivating a deep critical awareness of how systems operate or commodify data. Participation remains mainly symbolic, with marginalized groups, such as older adults or those lacking digital fluency, frequently excluded from the benefits of innovative city initiatives.

African cities face additional challenges in pursuing equitable development of AI. High hardware costs, reliance on foreign cloud infrastructure, and significant digital literacy disparities hinder the development of local, independent AI ecosystems. Furthermore, dominant narratives of "smartness" often reflect external ideals and fail to consider the lived experiences and knowledge systems of local communities. AI does not arrive in these contexts on a blank slate; it is interpreted, adapted, and reimagined through existing cultural practices and knowledge systems. It reminds us that each technological innovation also constitutes a disruption, as it often removes, alters, or renders obsolete previous tacit knowledge and established practices in the pursuit of innovation.

In all contexts, the growing presence of AI has introduced a new landscape of legal, ethical, social, and technological uncertainty. Even within municipal governments, the adoption of AI remains limited and experimental. Many proposed applications are still in the pilot stage, with their benefits yet to be fully understood, let alone proven, while their operational challenges and side effects continue to accumulate.

Incidents linked to issues with the development, use, and governance of AI misuse are on the rise. These range from data breaches to unapproved deployments, underscoring the urgent need for anticipatory reflections and accountability. Moreover, we need to stay aware of the gap between ethical aspirations and legal enforceability. While early frameworks, including the EU's Trustworthy AI Guidelines, were centered on sustainability, these ideals have been diluted in the hard law provisions of the AI Act and international agreements. It is therefore important that governance structures can make use of the binding mechanisms needed to ensure meaningful environmental accountability, rather than focusing on voluntary codes, and in doing so make sustainability a core principle, not an afterthought.

Together, our book and our discussions advocate for a democratic, transparent, and human- and environmentally centered vision of AI governance. This requires ***stronger regulation***, such as the European Union's AI Act, as well as ***anticipatory foresight and real-time reporting mechanisms***. A human-centered approach to AI must move beyond viewing people as passive recipients of technological change and instead position them as co-creators, safeguarding the fundamental human aspects of labor, care, and creativity. A human-centered vision can only be sustainable if this approach aligns with the needs of our planet, which in turn supports the well-being of humanity as a whole. This necessitates moving beyond a narrow, individualistic interpretation of human-centeredness to also consider the interests of those people typically not seen or heard, the needs of diverse communities, the complexities of dynamic societies, and the planet as an integrated system.

As such, we believe that ***transparency*** is essential. The opaque workings of AI, its data sources, training processes, and decision-making logic, must be made legible to the public. Only through this visibility can we begin to demand accountability for the social and ecological costs of AI systems. Ethical frameworks must be institutionalized to ensure that AI does not entrench power asymmetries or replicate existing injustices. Public education should focus not only on technical skills but also on critical digital literacy, empowering individuals to understand, question, and shape the technologies they use.

Moreover, the importance of ***social dialogue*** cannot be overstated. For AI to be truly democratic, the communities affected must be involved in designing, deploying, and governing it. This involves establishing participatory structures in which workers, citizens, and policymakers can deliberate on the future of AI, proactively shaping it rather than simply reacting to change. Finally, all AI development must be underpinned by a commitment to environmental justice. Recognizing the material realities of AI, including its energy demands, extractive supply chains, and carbon footprint, requires a cultural shift toward slower, more intentional forms of innovation, rather than *moving fast and breaking things*.

As Alan Kay once stated, "The best way to predict the future is to invent it."[7]

7 Kay, A. C. (1989). *Predicting the future*. Stanford Engineering, *1*(1), 1–6.

Ultimately, the future of AI is not fixed. It remains open to contestation, co-creation, and care. The decisions we make now about how AI is built, used, and governed will shape its social contract for generations. Instead of succumbing to the allure of inevitability, we must take on a shared responsibility to envision and implement an equitable and sustainable AI future that serves the common good.

Authors

Aída Ponce Del Castillo

Aída Ponce Del Castillo holds a PhD in law (Doctor Europaeus) and a master's degree in bioethics. At the European Trade Union Institute's (ETUI) Foresight Unit she conducts research on the ethical, social, and legal implications of emerging science and technology, with a specific focus on AI, and coordinates the institute's foresight projects. She sits on Belgium's Committee for the National Convergence Plan for the Development of AI and represents the ETUI in OECD working parties on biotechnology, nanotechnology, and converging technologies, as well as AI governance. Previously, she headed the ETUI Health and Safety Unit and coordinated the Workers' Interest Group at the Advisory Committee on Safety and Health to the European Commission.

Ans De Nolf

Ans De Nolf is a doctoral researcher at KU Leuven, specializing in media, social justice, and emerging technologies. Her work focuses on media representations, Islamophobia, youth and digital resilience, and the socio-ecological implications of AI. She has contributed to several international research projects, including DEMINE, ySKILLS, and AI(m) for the Future, and is a member of the KU Leuven Digital Society Institute (DIGISOC). Her publications include work in *Journalism and Media*, *RELIGIONS*, and edited volumes with Routledge and Leuven University Press. Through her interdisciplinary and participatory approach, she seeks to contribute to more inclusive, ethically grounded, and socially responsive forms of media and technology.

Benedetta Brevini

Benedetta Brevini is Associate Professor of Political Economy of Communication at the University of Sydney, Visiting Professor at the Institute for Public Knowledge at New York University, and Senior Fellow at the London School of Economics. Her research explores the intersections of communication, data capitalism, and the climate emergency. She is the author of several influential books, including *Is AI good for the planet?* (2021), which was named *Nature*'s science book of the week. In addition to her academic work, Brevini contributes regularly to outlets such as *The Guardian*, *South China Morning Post*, *OpenDemocracy*, and *The Conversation*. She is currently working on a new book for Polity Press.

Bettina Berendt

Bettina Berendt is Professor for Internet and Society at the Technical University of Berlin since 2019, Director of the Weizenbaum Institute, and Guest Professor at KU Leuven, Belgium. Her current research includes data science and critical data science, especially with respect to privacy/data protection, discrimination, and fairness, as well as AI and ethics, with a focus on textual and web-related data. Until 2019 Bettina Berendt was professor at the Department of Computer Science / Research Group Declarative Languages and Artificial Intelligence at the KU Leuven and before that junior professor at the Institute of Business Informatics at the Humboldt University Berlin. Details on publications, lectures, projects, positions, courses, and other activities can be found at www.berendt.de/bettina.

Bieke Zaman

Bieke Zaman is Professor of Communication Sciences and Human–Computer Interaction at KU Leuven, where she leads the Meaningful Interactions Lab (Mintlab) and serves as the founding director of the KU Leuven Digital Society Institute (DigiSoc) since 2021. Her research explores how children, youth, and broader communities engage meaningfully with digital technologies through participatory design, art–science interactions, and mixed methods. Zaman investigates converging domains such as gaming, gambling, and sports, and champions creative avenues for science communication and co-creation. She teaches courses in HCI, media design, and qualitative methods, and holds a PhD in social sciences from KU Leuven (2011).

Bruno Dupont

Bruno Dupont is the Institute Coordinator of the interdisciplinary KU Leuven Digital Society Institute. At KU Leuven, he is also a postdoctoral research manager and postdoctoral researcher in Media Studies. He was responsible for the daily management of the FWO-SBO project Gam(e)(a)ble, led by Rozane De Cock and Bieke Zaman. Within this project, Bruno studied the intersection between gaming and gambling in contemporary digital media from a media studies angle, in connection with a team of psychologists, legal researchers, and prevention specialists. He is particularly interested in the development of frameworks for media education and ludoliteracy, both aimed at future game developers and the wide audience.
Next to Gam(e)(a)ble, Bruno remains true to his background in literary and game studies, and researches digital literature, emergent media forms, and intermedia phenomena in literature. Outside of KU Leuven, Bruno was also Invited Professor and Research Coordinator at the Brussels Institute for Teacher Education (VUB).

Chao Wang

Chao Wang is a joint PhD candidate at Sun Yat-sen University (China) and KU Leuven (Belgium), affiliated with the Media Culture & Policy Lab and the Digital Society Institute (DigiSoc). His research explores media and information literacy, digital and AI literacy, and smart libraries. Chao is a Standing Committee Member of the IFLA Information Literacy Section and an expert with the UNESCO MIL Alliance. He has presented internationally and published widely in the field.

Elisabeth Struyf

Elisabeth Struyf is a communications educator and program coordinator at KU Leuven, where she supports the Master of Communication Sciences: Digital Media and Society. With a background in education policy and digital media, she bridges academic learning and real-world experience. She co-teaches courses on digitalization and plays a central role in organizing student internships, workshops, and hackathons. Passionate about innovative teaching and student development, Elisabeth works closely with both faculty and industry partners to help students explore the evolving role of media and technology in society.

Florence Chee

Florence Chee is Associate Professor for the School of Communication at Loyola University Chicago, where she directs the Center for Digital Ethics and Policy and founded the Social & Interactive Media Lab Chicago (SIMLab). Her research focuses on the social, cultural, and ethical dimensions of digital technologies, including AI, games, social media, and mobile platforms. She engages across academic, industry, and policy sectors to inform responsible innovation. Dr. Chee serves as an External Consultee to the Freedom Online Coalition's Taskforce on AI and Human Rights and is a Key Constituent of the UN 3C Roundtable on AI.

Frank Pasquale

Frank Pasquale is Professor of Law at Cornell Tech and Cornell Law School. He is an expert on the law of AI, algorithms, and machine learning. His books include *The black box society* (Harvard University Press, 2015) and *New laws of robotics* (Harvard University Press, 2020). He has published more than 70 journal articles and book chapters, and co-edited *The Oxford handbook on the ethics of artificial intelligence* (Oxford University Press, 2020).

Joep Crompvoets

Joep Crompvoets is Professor and Research Manager at the KU Leuven Public Governance Institute, where he holds the Chair in Information Management in the Public Sector. He is also the founder and scientific coordinator of the Erasmus Mundus Joint Master's program Public Sector Innovation and E-Governance, a collaboration between KU Leuven, the University of Münster, and Tallinn University of Technology. His research and teaching focus on spatial data infrastructures, data science, geographic information systems (GIS), public sector innovation, digital transformation, interoperability, and e-governance.

Kamile Grusauskaite

Kamile Grusauskaite is an FWO-funded postdoctoral researcher at KU Leuven's Media Culture & Policy Lab and a visiting fellow at Yale University's Center for Cultural Sociology (2024–2025). A cultural sociologist with a PhD from KU Leuven, her work examines disinformation, conspiracy theories, platform governance, and digital communities through qualitative and computational methods. Using ethnography, digital observation, and interviews, she explores how creators shape narratives, resist deplatforming, and navigate infrastructural constraints. Her research appears in *Cultural Sociology*, *New Media & Society*, *Social Media + Society*, *Public Understanding of Science*, and more.

Katja Biedenkopf

Katja Biedenkopf is Professor of Sustainability Politics at KU Leuven's Leuven International and European Studies (LINES), where she leads the Sustainable Futures research group. Her work focuses on climate and environmental governance, especially climate diplomacy, policy learning, diffusion, and leadership, as well as polycentric governance. Biedenkopf teaches courses on European and global sustainability policy and politics through interactive methods.

Katrien Pype

Katrien Pype is a social and cultural anthropologist specializing in urban life, media, and technology in central Africa. Since 2003, she has conducted extensive ethnographic research in Kinshasa, the capital of the Democratic Republic of Congo. Her work explores popular culture, religious expression, and technological innovation.

Her monograph, *The making of the Pentecostal melodrama: Religion, media and gender* (Berghahn Books, 2012), examines the vibrant media landscapes shaped by Pentecostalism. Her research has been published in leading journals such as *Media, Culture & Society*; *Africa: Journal of the International Africa Institute*; *Ethnos: Journal of Anthropology; Journal of the Royal Anthropological Institute; Anthropological Quarterly;* and *City & Society*. She co-edited *Cryptopolitics: Exposure, concealment, and digital media* (2023, with V. Bernal and D. Rodima-Taylor) and the forthcoming *The post-global city: Theorizing technology cultures in urban Africa* (2026, with O. Adunbi and M. M. J. Fischer). She promotes research on religion, tech creativity, and climate change across Africa and beyond.

Kristof Vrancken

Kristof Vrancken is a lecturer, researcher, and visual artist in Media Culture & Policy at KU Leuven's Faculty of Social Sciences and LUCA School of Arts, where he coordinates the research cluster and team "Future Scenarios". He is a member of DigiSoc, the university's Digital Society Institute. His research engages with complex environmental issues, regenerative futures, and societal challenges, such as dementia and migration, through participatory and artistic research methods. His work has been published in international peer-reviewed journals and books, and presented at international scholarly conferences. His artistic practice and research have been featured in exhibitions at prominent institutions both in Belgium and abroad.

Kurt Feyaerts

Kurt Feyaerts is Professor of Linguistics at KU Leuven and a member of both the Urban Studies Institute (LUSI) and the Leuven Interdisciplinary Language Institute (LILI). He co-founded the MIDI research group (Multimodality, Interaction & Discourse) and teaches courses on (German) linguistics, humor, creativity, and multimodality across Leuven and Kortrijk campuses. At Venice International University, he teaches and coordinates courses and summer schools on multimodal ecolinguistics and urban linguistic landscapes. His research focuses on the multimodal dynamics of meaning-making, exploring how language, space, and materiality interact as communicative resources in shaping urban environments and intersubjective understanding.

Leen d'Haenens

Leen d'Haenens is Professor of Communication Science at KU Leuven, where she chairs the Media Culture & Policy Lab. Trained in languages, communication, and information studies (Ghent, Toronto), she holds a PhD in political and social sciences from Ghent University. She studies topics such as media framing of migration and Europe, digital literacy among youth, and civic engagement among marginalized communities. She follows a cross-national, interdisciplinary approach and has led or contributed to major international projects, including ySKILLS and EU Kids Online. She is editor-in-chief of *Communications* and co-editor of the *International Communication Gazette*. At KU Leuven's Digital Society Institute (DigiSoc), she leads the "Democracy and Civic Engagement" pillar, promoting research at the intersection of media, technology, and society.

Meryem Soyturk

Meryem Soyturk is a journalist and digital media researcher from Türkiye, currently pursuing a master's degree in Communication Sciences: Digital Media and Society at KU Leuven. With over six years of newsroom experience, she has covered breaking news and led social media strategy. Her academic focus centers on the integration of AI in journalism, exploring its impact on newsroom efficiency, ethics, and labor. Meryem blends her professional background with academic inquiry to explore the evolving media landscape. She is especially interested in global media innovation, responsible technology use, and the future of journalism in an AI-driven era.

Peggy Valcke

Peggy Valcke is Executive Board Member at BIPT, where she leads the Legal Service, HR and NetSec departments. She is also part-time Professor of Law & ICT at KU Leuven, co-director of imec-CiTiP, and member of Leuven.AI. Her main areas of expertise cover media and communications, data protection and data governance, cybersecurity, and AI. She is head of the Belgian delegation at CAI, the Council of Europe's Committee on AI that prepared and negotiated the Framework Convention on AI, and previously served as Vice-Chair of CAI's predecessor, CAHAI. She was one of the founding co-directors of the Flemish Knowledge Centre Data & Society until she joined BIPT in January 2024.

Pepijn Viaene

Pepijn Viaene holds a Master of Arts in philosophy (KU Leuven) and completed supplementary studies in social and cultural anthropology after a year of fieldwork and volunteer work in Kinshasa (2000–2001). Since 2023, he has been a research fellow at the KU Leuven Lab for Media Culture & Policy. His research focuses on digital well-being and the use of AI-driven apps and technologies by children with autism in classroom settings in both Kinshasa and Flanders. He resumed ethnographic fieldwork with two two-week stays in Kinshasa in 2023 and 2024, as preparation for a future PhD project. Initially exploring AI in the workplace, he now examines how such tools can support neurodivergent learners, bridging technological innovation and inclusive education.

Sandy Claes

Sandy Claes is a part-time assistant professor at the lab for Media Culture & Policy of KU Leuven, which she combines with a faculty position at LUCA School of Arts, Belgium. In her research, she focuses on the overlap between media, technology, and public space, which she studies via a research-through-design approach. Current research projects include representing uncertainty in data visualization for news and science, and facilitating civic participation through in-situ media. Previously, she has worked as a lead user researcher at the innovation department of public broadcaster VRT. She received her PhD in engineering science (2017) from KU Leuven with her work on public visualization. She has a background as a master in audiovisual arts (2005). Her audiovisual work has been awarded at several international film festivals; such as I Castelli Animati in Rome and the international short film festival of Leuven, and has been exhibited at several international venues; such as LABoral, Madrid, and Museum M, Leuven. With this mixed background of science, design, and arts, Sandy approaches research projects from a multidisciplinary viewpoint.

Sercan Kiyak

Sercan Kiyak is a researcher preparing a PhD at the Media, Culture & Policy Lab at KU Leuven, where he investigates political communication on social media, with a particular focus on migration discourse. His research integrates social network analysis, topic modeling, and digital methods to examine public opinion, polarization, and framing across digital platforms. Kiyak has contributed to international conferences and peer-reviewed publications on topics such as the refugee crisis and climate-induced migration. He was previously involved in interdisciplinary collaborations, including the EU-funded OPPORTUNITIES project. His work lies at the intersection of media sociology, natural language processing, and networked public discourse.

Tania Azadi

Tania Azadi (PhD) is a postdoctoral researcher at the Media Culture & Policy Lab, KU Leuven. Her work explores media and information literacy, focusing on vulnerable groups through participatory methods and application of critical theories. During her postdoctoral fellowship, Tania has been instrumental in developing and strengthening the Creative Europe Programme research project, PRODIGI, in which she also holds a co-leadership role. She received the 2022 and 2024 Marie Skłodowska-Curie Actions Seal of Excellence and the 2016 Erasmus Mundus MARHABA grant. Tania has taught courses in health and nursing informatics, medical information systems, and research methodology. She has translated key texts into Persian and serves as a peer reviewer for journals including *Journal of Immigrant & Refugee Studies*, *International Communication Gazette*, and *Communications*. She has translated key texts into Persian and serves as a peer reviewer for academic journals in communication sciences. Her interests include media and information literacy, information evaluation behavior, mis/disinformation, and community network engagement.

Stefan Mertens

Stefan Mertens is a postdoctoral researcher working on issues of ethnicity, diversity, and ideology across various research projects in Flanders and beyond. He is currently affiliated with PXL-MAD: Media, Arts & Design in Hasselt and the Department of Movement Sciences at KU Leuven. His work combines quantitative and qualitative methods to explore questions of intergroup relations, cultural identity, and public discourse. His research interests include survey research, discourse analysis, diversity studies, and the intersections of media, culture, and ideology.

Valeria Pulignano

Valeria Pulignano is Professor of Sociology and Francqui Research Professor at KU Leuven. Her research focuses on work, employment relations, and labor markets, with particular attention to inequality, precarity, and workers' voice. She coordinates the Work, Employment and Industrial Relations network at the European Sociological Association and is affiliated with CRIMT, Warwick University, and LISER. She serves as Editor of Work, Employment and Society and has published widely, including books with Oxford University Press and Palgrave. Her forthcoming book with OUP explores how unpaid labor shapes inequality in precarious work.

Vangelis Palaskas

Vangelis (Evangelos) Palaskas is an EU and digital policy consultant specializing in AI and data governance. With certifications in PM², CSPO®, and CSM®, he brings a structured and agile approach to policy development and project management. Vangelis is currently pursuing a PhD focused on the European Union's digital transition, combining academic research with hands-on policy expertise to support effective and ethical digital transformation in Europe.

Zeynep Hamurdan

Zeynep Hamurdan is a communications professional and researcher specializing in digital media, EU policy, and political communication. She served for over ten years in the Turkish public sector as a communications officer. Currently, she is pursuing a Master of Science in Digital Media and Society at KU Leuven as a Jean Monnet Scholar. She also holds a master's degree in mass communications from Southern Illinois University Edwardsville, where she was awarded a Research Grant for Graduate Students. Her academic journey began with a bachelor's degree in political science and public administration from Hacettepe University, Türkiye. Throughout her academic and professional career, she has focused on the intersection of media, governance, and technology.

AI: RETHINK, REGULATE, REIMAGINE

What happens when AI technologies manage workers instead of human employers? Can AI help fight climate change while also fuelling new forms of environmental exploitation? Will "smart cities" empower citizens, or silently erode civic trust? This book captures the AI Dialogues held during KU Leuven's 600th anniversary, where scholars, practitioners, and policymakers from across Europe, the U.S., Australia, China, and Africa explored how algorithms shape labour, sustainability, and urban life. Through lectures, working sections, and debates, the opportunities and risks of AI were discussed: from reshaping employment to the ecological costs. The volume brings together these diverse perspectives in a shared call for democratic, ethical, and sustainable approaches to AI—ensuring innovation truly serves the public good.

ABOUT THE AUTHORS

Leen d'Haenens is Professor of Communication Sciences at KU Leuven, where she chairs the Media Culture & Policy Lab and leads the "Democracy and Civic Engagement" pillar at the Digital Society Institute. Her research focuses on migration framing, media diversity and pluralism, and the development of media literacy and digital skills.

Ans De Nolf is a researcher at the Media Culture & Policy Lab at KU Leuven and the author of a doctoral dissertation on Islamophobia, Muslim youth and resilience on- and offline. Her work focuses on media representations, social justice, and civic engagement.

Bieke Zaman is Professor of Communication Sciences at KU Leuven, where she leads the Meaningful Interactions Lab (Mintlab) and directs the Digital Society Institute (DigiSoc). Her research focuses on how children, youth, and communities engage with digital technologies, spanning domains such as gaming, gambling, sports, and participatory design.

acco learn