Coll. Antropol. **49** (2025) 1: 7–13
Original scientific paper
doi:10.5671/ca.49.1.2

AI for Social Good Begins with Accountability

Rahul Dodhia

Microsoft AI for Good Lab, Seattle, USA

ABSTRACT

Artificial intelligence is often framed through questions of benefit and risk, yet its effects arise from how it reshapes the conditions under which knowledge is produced and acted upon. This paper examines AI for social good as a set of practices that extend the capacity to observe, classify, and infer at scales beyond human perception. These systems change what becomes visible, how that visibility is interpreted, and who is positioned to make use of it. As machine inference assumes more of the observational and interpretive load, authority shifts toward the institutions that design and govern these tools. The analysis identifies accountability, transparency, and equity as the practical conditions that determine whether AI strengthens or weakens the communities and environments it touches. The aim is to clarify how AI reorganizes the relationship between observation, judgment, and responsibility within projects framed as socially beneficial.

Key words: *AI for social good, AI accountability and transparency, machine learning, AI governance, responsible AI*

Introduction

Public unease about artificial intelligence resembles moments in history when societies confronted far-reaching technologies[1]; technologies so universal that they fall under the term general purpose technologies (GPTs). People often place greater weight on what might be lost when the familiar gives way to the new, a tendency shaped by well-documented cognitive inclinations that favor stability over uncertainty[2,3]. These reactions surfaced around steam power, internal combustion, electrification, and digital computing, each of which arrived with strong expectations of progress and a parallel sense of risk.

AI is the latest GPT entry, but where it diverges from earlier GPTs is in its direct extension of human cognitive processes. Previous GPTs augmented perceptual reach or mechanical capability. AI, by contrast, operates on the level of inference and reasoning, effectively externalizing functions once confined to the human mind[4,5]. For the first time, technological development is directed toward modelling and amplifying the processes of cognition itself. Systems that predict protein structures, draft legal arguments, or allocate medical attention work in domains previously reserved for human judgment. These capabilities change how knowledge is produced and who is positioned to produce it, and they reshape the conditions under which people and communities become visible to institutions[6].

Because of this, the question should not be whether AI inspires fear or enthusiasm. What matters is how AI reconfigures visibility, authority, and the practical conditions under which people receive care, assistance, or recognition. A village in Niger becomes legible on a map that once omitted it; a premature infant receives timely screening despite the absence of local specialists; a conservation team detects species declines that would otherwise go unnoticed. The same systems can expose dissidents, misclassify civilians, or render whole communities invisible when they fall outside the categories a model relies on[7].

This paper examines these dynamics across three domains where AI already shapes decisions with material consequences: humanitarian response, environment and conservation, and medicine. Each domain reveals how AI can widen access to information, extend human perception, and support forms of expertise that were previously out of reach. Each also shows how AI can concentrate power, narrow forms of recognition, or diffuse responsibility. The analysis then turns to accountability, transparency, and equity as the central conditions that determine whether AI systems operate as tools of support or tools of harm.

The goal is not to resolve the debate over whether AI is beneficial or dangerous. A different framing might prove more actionable. AI amplifies the intentions and structures of the societies into which it is introduced. Understanding how it is used and how it can be governed

requires attention to the ways it recasts what is seen, who is authorized to act on what is seen, and how those decisions are explained or can be contested.

AI for Social Good

Humanitarian response

Disasters are not just physical events; they are also crises of information. When floods, earthquakes, or storms strike, the first question is never only "How many are dead?" but "Where are they? Who is still alive? Who has been forgotten?" Disaster response is a race against time and urgent need for information, one in which the availability and interpretation of data often determine survival.

In the 2023 Derna floods in Libya, two dams collapsed before dawn and erased entire districts[8] . Communications went dark. Rescue teams could not tell which roads still existed, which bridges were standing, or where the living and dead lay. In previous decades, it would take days, optimistically, for aerial photographs to be collected, manually interpreted, and converted into usable maps.

With AI capable of mimicking human perception and cognition, satellite images taken hours after the disaster were processed automatically to detect destroyed buildings. Damage assessments that previously took days were completed overnight[9]. These kinds of capabilities could transform how first responders operate. Search teams no longer have to choose blindly where to go first; they can follow updated maps, often created faster than rescue vehicles could reach the site. AI's role in this case is not to supplant human judgment, but to compensate for the limits of human perception, which cannot scan millions of pixels across thousands of square kilometers within a life-saving timeframe.

Beyond improving efficiency at vast scales, AI helps make visible that which was unseen before. In Zinder, Niger, one of the poorest regions in the world, entire villages had never appeared on an official map. Without census data, aid could be distributed based on outdated national statistics, not local realities. Using satellite imagery and AI models trained to detect rooftops, field huts, and settlement patterns, population density maps can be produced that reveal communities previously unrecorded by the state.[10]

Anthropologically, visibility confers political and social existence: one becomes eligible to be helped, vaccinated, educated, or governed.[11] AI did not create these populations; it simply made it harder for systems to ignore them. Yet visibility cuts both ways. A satellite that finds flood survivors can also find refugees crossing a border. A drone that identifies collapsed buildings can be used to monitor protests or track dissidents. The same tools that make people visible to rescuers can make them visible to those who would harm or control them.

In some humanitarian contexts, displaced communities live in camps that lack up-to-date maps or infrastructure records. A recent project covering the Kakuma–Kalobeyei refugee camps in Kenya used high-resolution aerial imagery combined with machine learning to produce detailed maps of shelters, roof materials, sanitation facilities, and energy infrastructure, types of data rarely available for refugee settlements. By making formerly unmapped dwellings visible to humanitarian organizations, these AI-driven maps enable resource planning, aid distribution, and infrastructure interventions in ways that traditional remote-sensing or manual survey methods could not match[12].

This redistribution of "vision" allows actors beyond large international agencies, such as local NGOs, camp residents, humanitarian workers, to see, act, and advocate for communities that might otherwise remain invisible[13]. Visibility, in this sense, is empowerment: it allows more actors to participate in the response, to contest official narratives, and to make claims on the state[13]. The same systems that open up visibility can also shape what is seen. The engineers designing these models determine what the AI is trained to see by choosing the categories it recognizes, the data it ingests, and the thresholds it deems significant. Some patterns of life may remain invisible but because they were never encoded as data in the first place, unintentionally or otherwise.

Environment, climate, and conservation

Before modern meteorology, farmers and pastoralists forecast rain through long familiarity with local conditions. These practices were grounded in lived environments and generational knowledge. However, as climate change accelerates, familiar cues have become less reliable, and technology has had to step in. Satellites scan the Earth using high resolution sensors, and generate terabytes of data, often available for free. The constraint is no longer information but the ability to make sense of it before crops fail or species decline.

Agriculture

During the war in Ukraine, traditional crop-monitoring systems collapsed and no official reporting was possible in many regions. AI models trained on Sentinel-2 imagery classified wheat, barley, and maize across millions of hectares and predicted yields even where surveyors could not go[14]. AI systems like this certainly accelerate analysis, but they also shift how agricultural knowledge is produced. inference once made through field visits or national reporting now comes from satellites and automated models. The capacity to see the harvest and therefore anticipate scarcity moves toward those who control satellite data and the models that interpret it.

Energy

AI also redraws visibility in the global energy transition. Through initiatives like Global Renewables Watch[15],

machine learning models identify solar panels and wind turbines worldwide with a granularity that previously required costly site surveys. Now, environmental progress can be verified against physical installations rather than official statements, redistributing the ability to check compliance with climate commitments

Yet remote visibility can shift interpretive control away from people who live in the affected area[16]. When damage assessments are produced through satellite classifications or other automated analyses, the categories of what counts as destroyed, at risk, or in need are defined by remote analysts and institutions rather than by local actors[17]. Empirical studies of satellite-based assessment and digital humanitarian mapping show that these tools concentrate expertise in a limited set of organizations and that their outputs sometimes overlook informal settlements, livelihood spaces, or social ties that matter for response[13]. In such cases, local knowledge is not absent, but it enters decision-making on unequal terms, because remotely produced maps and indicators often carry greater authority in discussions about where to send resources and how to describe conditions on the ground.

Conservation

In biodiversity conservation, AI transforms how ecosystems are monitored. Rainforest soundscapes contain tens of thousands of hours of audio monthly, and camera traps produce millions of images. AI models detect elephant rumbles, rare bird calls, or the faces of chimpanzees in real time. These are tasks no team of researchers could perform manually[18].

As AI becomes embedded in ecological monitoring, it shapes which environmental changes draw attention and how those changes are judged. Patterns that align with easily obtainable data may come to define assessments of decline or recovery[19], while information known locally but absent from formal datasets can lose influence[20]. This shift places greater interpretive weight in the hands of institutions that manage large datasets and inference systems, even though they may be very distant, geographically and culturally, from the ecosystems in question. Conservationism must therefore account for how these judgments are formed and ensure that machine inference remains connected to the knowledge held by the communities who live within the ecosystems being monitored.

Medicine and Biology

Clinical practice of medicine depends a social contract of trust, care, and responsibility. AI enters this space promising new forms of diagnosis and, even more fundamentally, a new way of knowing the body. AI is also reshaping the biological sciences that inform medicine through models that are experiencing spectacular triumphs, such as in protein folding, radiology, and pathology. Here, questions of who benefits, who is accountable for AI's outputs, and how decisions are can be made legible to doctors and patients, loom large.

Protein folding

For decades, predicting how a protein folds from a linear amino acid sequence into a functional three-dimensional form was a central challenge in biology, the holy grail of modern biology.[21] The way proteins fold determines how the body works: how viruses bind, how enzymes repair DNA, how drugs function. Experimental methods like X-ray crystallography and cryo-electron microscopy could solve one protein in months or years, with tens or hundreds of thousands more to go.

In 2020, AlphaFold solved over 200 million protein structures within a few months. It did not use physical measurement and theoretical physics or chemistry; it used pattern recognition. AI "learned" the rules of life from past data, then predicted what biology had not yet measured.[22] AI is changing who can contribute to discovery. Students without access to traditional laboratories can now pursue protein design computationally, and researchers in regions without advanced experimental infrastructure can participate in early-stage drug development. For much of modern science, authority has rested with those who controlled specialized instruments and well-funded institutions. Now, the growing reliance on computational models shifts part of that authority toward those who develop, maintain, and interpret these systems. The well-funded laboratory is still required to validate predictions, but the first steps of inquiry increasingly originate on model servers rather than in physical labs. In this context, AI functions as a new expert in the workflow, while responsibility for its outputs remains with the people and institutions that build and deploy it.

AI in healthcare

Retinopathy of Prematurity (ROP) can blind infants born too early. It disproportionately affects babies weighing under 1.06 kg[23], especially in regions where neonatal survival has improved faster than specialist care. As of 2012, there were only 200,000 ophthalmologists worldwide, and a fraction were pediatric.[24] There are just not enough doctors to fulfil the demand.

In South and Central America, health practitioners can now use smartphones with retinal lenses to photograph premature infants' eyes. AI models flag cases needing urgent treatment. Images can be sent to specialists hundreds of kilometers away. AI does not replace doctors, it creates medical attention where none existed.

Healing often involves more than biomedical assessment; it includes communication, reassurance, and shared interpretation between patients and clinicians. AI enters this arrangement as a new subject-matter expert offering judgments expressed as risk scores or classifications, and potentially even as verbal explanations. However, they do not fully reveal how the underlying statistical system reached its conclusions. Families cannot examine the in-

ternal basis of a model's decision, and clinicians may be unsure whether a generated explanation reflects genuine reasoning or a post-hoc description. The black box of algorithmic inference modifies the social structure of care: Accountability now depends on data quality, calibration procedures, and institutional oversight. Responsibility becomes distributed across the model, the organization that deploys it, and the datasets that shape its behavior. These changes do not displace clinical relationships, but they introduce new forms of mediation that must be managed so AI supports rather than complicates the moral and practical work of care.

This is not new. Medicine has long translated bodies into data through temperature charts, laboratory values, imaging studies, and other measurements designed to reveal patterns beyond individual perception. AI amplifies this, and what distinguishes AI is its ability to generate visibility at scale. Millions of images and signals can be interpreted in seconds, revealing pathologies or trends no clinician could detect unaided.

This new visibility also reshapes power and expertise. Diagnostic authority used to be concentrated in hospitals and specialists and now extends to field workers equipped with mobile phones and portable sensors. In rural clinics, an image captured by a nurse can trigger an AI assessment and a specialist's response within minutes. Expertise thus becomes distributed more widely as a collaboration between human experience and algorithmic perception. However, the asymmetry of power remains uneven: the designers or owners of the models still determine which diseases are detectable, which bodies are represented, and which remain unseen.

Who Takes Responsibility for AI

AI does not set its own aims. Its behavior reflects the data it is trained on and the objectives defined for it. These systems can support careful analysis or amplify errors, depending on how they are designed and governed. As AI becomes more common in governance, science, and humanitarian work, responsibility shifts toward the institutions that assemble the datasets and specify the conditions under which models operate. This section examines three conditions that influence how the authority these systems acquire translates into responsibility: accountability, transparency, and equity.

Accountability

AI systems promise to enhance decision-making, but it's not always clear who is answerable when harm occurs[25]. In computer-vision applications, artificial neural networks can identify collapsed infrastructure after floods or earthquakes, producing maps that guide humanitarian logistics within hours. These tools extend institutional capability and save lives. However, the same computing architectures can be used by surveillance technology that continuously track populations through facial recognition or aerial imagery. When automated systems perform tasks previously carried out by identifiable professionals, the chain of responsibility becomes harder to trace.

A clear example is the 2020 wrongful arrest of Robert Williams in Detroit: a facial-recognition system misidentified him, police treated the algorithmic output as evidence, and subsequent inquiries showed that neither the software vendor nor the officers involved accepted responsibility for the error. When an AI-generated map misclassifies civilian structures as military targets, or a facial-recognition system leads to wrongful arrest[26], accountability is frequently displaced among developers, vendors, and governments. This diffusion of responsibility contrasts sharply with traditional professional ethics, where identifiable human agents, such as pilots, physicians, engineers, are accountable for outcomes. When the chains of responsibility are vague, model decisions can circulate without an identified person or institution who can explain or defend them.

Legal scholarship and science-and-technology studies converge on a central requirement: every high-risk AI system must have a clearly identifiable human or institutional agent answerable for its outcomes[27].

Just as aviation and pharmaceutical industries evolved liability regimes in keeping with their complexity[28] , AI governance must develop comparable standards of due care.

Concrete mechanisms should include:

- legally mandated human-in-the-loop oversight for consequential decisions (e.g., healthcare, finances, border control).
- independent audit authorities empowered to inspect training data and model performance.
- procedural rights for individuals to obtain explanation and appeal of algorithmic decisions.

Accountability also has a cultural dimension. As Peacock et al. (2021) note, visibility entails moral obligation: those who render others legible assume a duty of care. The expansion of seeing therefore requires a parallel expansion of governance mechanisms, such as auditable datasets, public oversight boards, and avenues for redress, so that the capacity to perceive does not exceed the capacity to be held accountable.

Transparency

Predictive and generative models operate through complex statistical inference that is often hard to explain even by their designers[29]. In predictive policing, for instance, algorithms trained on historical arrest data may label certain neighborhoods as "high risk," prompting intensified patrols and further arrests, creating a self-reinforcing feedback loop[30]. The absence of interpretability transforms these models into epistemic authorities whose judgments appear objective but reproduce historical bias.

Large-language models present a similar challenge. They can produce text that resembles official guidance, emergency bulletins, or journalistic reporting, yet they

provide no clear account of how they arrived at these statements. Recent studies show that such systems can generate persuasive falsehoods that pass as credible information and are harder to identify than earlier forms of misinformation[31].

From a societal standpoint, opacity reconfigures the social contract of knowledge. People expect explanations from professionals whose decisions affect their lives, such as doctors, engineers, officials. When decisions arise from unexplainable computation, a form of accountability disappears[32]. Transparent AI must include not only open-source code, but explainable outputs, evidence-based sourcing, uncertainty disclosure, and pathways for human review and accountability. This might require allowing only models whose reasoning can be explained in language accessible to affected publics, auditable data provenance, and disclosure of uncertainty[33,34].

Equitable access

In many regions, limited availability of compute resources, internet and electric connectivity, and technical education prevent meaningful participation in building or adapting AI. This turns the digital divide[35] into something deeper, because the gap concerns the ability to influence the systems that shape economic life and public services. The benefits of AI will reach people across the world over time, yet the avenues for those benefits differ sharply. Weather prediction depends on networks of monitoring stations, which many countries lack[36]. Diagnostic models pivot on data from specific populations, usually from the richer, Western and Northern hemispheres[37], leaving them in doubt when applied to other populations; clinical data from many regions is scarce. Large models trained on Western text, imagery, and medical records absorb cultural assumptions about health, risk, and daily life[38,39]. These assumptions travel with the models and limit their fit outside the regions for which they were created.

Regions without strong computational capacity or regulatory leverage continue to rely on external platforms for data access and decision support. This dependence places them at a disadvantage in a core sector of the global economy. Within states, algorithmic systems can amplify unequal access to credit, jobs, and health care. Couldry and Mejias (2019) describe data extraction and algorithmic production as extensions of older colonial patterns, rewritten in contemporary technical language[40].

AI expands the reach of collective reasoning, yet it often reflects the inequalities of its origin. Those whose experiences shape the dominant datasets retain the strongest influence. Equity in AI governance requires attention to information, particularly about who appears in data and who is absent, and to material conditions, who carries the environmental and social weight of computation. Policies that broaden access to model capabilities, encourage efficient system design, and bring a wider range of people into dataset creation establish the groundwork for a fairer AI landscape[41].

Sustainability and global governance

Because AI is transnational, meaning it can be trained in one jurisdiction, deployed in another, and impactful everywhere, its regulation demands international coordination. Two areas illustrate this need. Autonomous weapons, made deadlier and cheaper by AI, demand treaties that ensure meaningful human control over lethal decisions to prevent destabilizing escalation. Cross-border data flows require agreements on ownership, consent, and equitable use, particularly for medical, geospatial, and social data that support globally deployed models. These issues reveal the broader fact that AI governance is inseparable from global political commitments[42].

Sustainability extends the need for governance to the physical substrate of AI. Training a large-language model can consume energy equivalent to that used by hundreds of households annually, and datacenter cooling competes with community water supplies. These impacts are unevenly distributed with the environmental cost borne where datacenters are sited, while the benefits of model deployment accrue globally. Because compute, energy, and water markets cross borders, the environmental footprint of AI cannot be treated as just a domestic matter. Coordinated standards for environmental reporting, lifecycle assessment, renewable-energy procurement, and resource management would bring AI into alignment with national climate targets and international sustainability commitments[43]. Treating AI as part of climate infrastructure rather than as an abstract digital service anchors model development within planetary limits.

Discussion

Across humanitarian response, environmental monitoring, and medicine, AI changes how the world becomes knowable. Systems that interpret satellite imagery, soundscapes, medical scans, or human speech alter the balance between direct experience and mediated inference. They extend the capacity to detect events, classify conditions, and anticipate risks, but they also shift where interpretive authority resides. What counts as evidence, and who is positioned to act on it, now depends on tools that operate at scales no person could sustain.

These shifts raise structural questions rather than technical ones. Visibility is uneven, and institutions with access to data and computational resources are often the ones who define what is seen. Communities that live with the consequences may have little influence over the categories and thresholds that guide these systems. AI's outputs can shape judgments that were once grounded in local knowledge, professional practice, or collective deliberation.

Aligning AI with social good therefore requires attention to the conditions under which its insights are produced. Accountability determines whether decisions prompted by models can be explained and challenged. Transparency clarifies how models classify or detect and what assumptions they embed. Equity ensures that data

infrastructures and inference systems do not reproduce longstanding exclusions. Sustainability places these tools within the broader ecological and material systems on which they depend. These are practical requirements, not abstract ideals.

AI will continue to expand the reach of human observation, but the value of that expansion depends on how its results are governed. The challenge is to ensure that the capacity to see does not eclipse the responsibility to understand, and that the authority to act remains connected to the lives and environments these systems are meant to support.

REFERENCES

1. EDGERTON D, The Shock of the Old: Technology and Global History Since 1900, (Oxford University Press, 2007). doi:10.1086/524257. — 2. BARBERIS NC, J Econ Perspect, 27 (2013) 173. doi:10.1257/jep.27.1.173. — 3. KAHNEMAN D, KNETSCH JL, THALER RH, J Econ Perspect, 5 (1991) 193. doi:10.1257/jep.5.1.193. — 4. HARARI YN, Nexus: A Brief History of Information Networks from the Stone Age to AI, (Random House Publishing Group, 2024). doi:10.1080/13518046.2025.2533628. — 5. SIEMENS G, MARMOLEJO-RAMOS F, GABRIEL F, MEDEIROS K, MARRONE R, JOKSIMOVIC S, DE LAAT M, Comput Educ Artif Intell, 3 (2022) 100107. doi:10.1016/j.caeai.2022.100107. — 6. TÖRNBERG P, SÖDERSTRÖM O, BARELLA J, GREYLING S, OLDFIELD S, Big Data Soc, 12 (2025) 20539517251338773. doi:10.1177/20539517251338773. — 7. DODHIA R, AI for Social Good: Using Artificial Intelligence to Save the World [Internet], (Wiley, 2024). — 8. UN OFFICE FOR THE COORDINATION OF HUMANITARIAN AFFAIRS, Libya Flood Response Flash Appeal Final Report Sept 2023 – June 2024 (OCHA, 2024). — 9. GHOLAMI S, ROBINSON C, ORTIZ A, YANG S, MARGUTTI J, BIRGE C, DODHIA R, FERRES JL, On the Deployment of Post-Disaster Building Damage Assessment Tools Using Satellite Imagery: A Deep Learning Approach. In: IEEE International Conference on Data Mining Workshops (ICDMW). 2022. doi:10.1109/ICDMW58026.2022.00134. — 10. GLAZER T, HACHEME GQ, ZAYTAR A, MAROTTI L, MICHAELS A, TADESSE GA, WHITE K, DODHIA R, ZOLLI A, BECKER-RESHEF I, FERRES JML, ROBINSON C, TEMPO: Global Temporal Building Density and Height Estimation from Satellite Imagery, (arXiv 2025). doi:10.48550/arXiv.2511.12104. — 11. SCOTT JC, Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed (Yale University Press, 1999). — 12. GUPTA A, ORTIZ A, NSUTEZO SF, KEBUT D, IYER S, DODHIA R, FERRES JML, Mapping Refugee Camps with AI: A Benchmark Dataset and Baseline Models for Humanitarian Applications. In: IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW, 2025). doi:10.1109/WACVW65960.2025.00059. — 13. BRYANT J, Digital Mapping and Inclusion in Humanitarian Response, (ODI, 2022). — 14. SADEH Y, WAGNER J, NAIR SS, OLIINYK O, BELLES E, BIBIH A, D'HARBOULLÉ L, BENDAHMANE H, GUPTA M, BECKER-RESHEF I, Sci Data, 12 (2025) 833. doi:10.1038/s41597-025-05190-7. — 15. ROBINSON C, ORTIZ A, KIM A, DODHIA R, ZOLLI A, NAGARAJU SK, OAKLEAF J, KIESECKER J, FERRES JML, Global Renewables Watch: A Temporal Dataset of Solar and Wind Energy Derived from Satellite Imagery [Internet], (arXiv, 2025). doi:10.48550/arXiv.2503.14860. — 16. BENNETT MM, Digit Geogr Soc, 8 (2025) 100116. doi:10.1016/j.diggeo.2025.100116. — 17. EDLER R, GLASZE G, KINZELBACH K, WALKER B, Front Remote Sens, 6 (2025). doi:10.3389/frsen.2025.1603575. — 18. MIAO Z, ELIZALDE B, DESHMUKH S, KITZES J, WANG H, DODHIA R, FERRES JL, Sci Rep, 15 (2025) 7242. doi:10.1038/s41598-025-89153-3. — 19. PRINGLE S, DALLIMER M, GODDARD MA, LE GOFF LK, HART E, LANGDALE SJ, FISHER JC, ABAD SA, ANCRENAZ M, ANGEOLETTO F, AUAT CHEEIN F, AUSTEN GE, BAILEY JJ, BALDOCK KCR, BANIN LF, BANKS-LEITE C, BARAU AS, BASHYAL R, BATES AJ, BICKNELL JE, BIELBY J, BOSILJ P, BUSH ER, BUTLER SJ, CARPENTER D, CLEMENTS CF, CULLY A, DAVIES KF, DEERE NJ, DODD M, DRINKWATER R, DRISCOLL DA, DUTILLEUX G, DYRMANN M, EDWARDS DP, FARHADINIA MS, FARUK A, FIELD R, FLETCHER RJ, FOSTER CW, FOX R, FRANCKSEN RM, FRANCO AMA, GAINSBURY AM, GARDNER CJ, GIORGI I, GRIFFITHS RA, HAMAZA S, HANHEIDE M, HAYWARD MW, HEDBLOM M, HELGASON T, HEON SP, HUGHES KA, HUNT ER, INGRAM DJ, JACKSON-MILLS G, JOWETT K, KEITT TH, KLOEPPER LN, KRAMERSCHADT S, LABISKO J, LABROSSE F, LAWSON J, LECOMTE N, DE LIMA RF, LITTLEWOOD NA, MARSHALL HH, MASALA GL, MASKELL LC, MATECHOU E, MAZZOLAI B, MCCONNELL A, MELBOURNE BA, MIRIYEV A, NANA ED, OSSOLA A, PAPWORTH S, PARR CL, PAYO-PAYO A, PERRY G, PETTORELLI N, PILLAY R, POTTS SG, PRENDERGAST-MILLER MT, QIE L, ROLLEY-PARNELL P, ROSSITER SJ, ROWCLIFFE M, RUMBLE H, SADLER JP, SANDOM CJ, SANYAL A, SCHRODT F, SETHI SS, SHABRANI A, SIDDALL R, SMITH SC, SNEP RPH, SOULSBURY CD, et al., Nat Ecol Evol, 9 (2025) 1031. doi:10.1038/s41559-025-02704-9. — 20. CIPRIANO C, NOCE S, MEREU S, SANTINI M, Ecol Model, 506 (2025) 111164. doi:10.1016/j.ecolmodel.2025.111164. — 21. BAEK M, DIMAIO F, ANISHCHENKO I, DAUPARAS J, OVCHINNIKOV S, LEE GR, WANG J, CONG Q, KINCH LN, SCHAEFFER RD, MILLÁN C, PARK H, ADAMS C, GLASSMAN CR, DEGIOVANNI A, PEREIRA JH, RODRIGUES AV, VAN DIJK AA, EBRECHT AC, OPPERMAN DJ, SAGMEISTER T, BUHLHELLER C, PAVKOV-KELLER T, RATHINASWAMY MK, DALWADI U, YIP CK, BURKE JE, GARCIA KC, GRISHIN NV, ADAMS PD, READ RJ, Baker D, Science, 373 (2021) 871. doi:10.1126/science.abj8754. — 22. JUMPER J, EVANS R, PRITZEL A, GREEN T, FIGURNOV M, RONNEBERGER O, TUNYASUVUNAKOOL K, BATES R, ŽÍDEK A, POTAPENKO A, BRIDGLAND A, MEYER C, KOHL SAA, BALLARD AJ, COWIE A, ROMERA-PAREDES B, NIKOLOV S, JAIN R, ADLER J, BACK T, PETERSEN S, REIMAN D, CLANCY E, ZIELINSKI M, STEINEGGER M, PACHOLSKA M, BERGHAMMER T, BODENSTEIN S, SILVER D, VINYALS O, SENIOR AW, KAVUKCUOGLU K, KOHLI P, HASSABIS D, Nature, 596 (2021) 583. doi:10.1038/s41586-021-03819-2. — 23. YUCEL OE, ERAYDIN B, NIYAZ L, TERZI O, BMC Ophthalmol, 22 (2022) 367. doi:10.1186/s12886-022-02591-9. — 24. RESNIKOFF S, FELCH W, GAUTHIER TM, SPIVEY B, Br J Ophthalmol, 96 (2012) 783. doi:10.1136/bjophthalmol-2011-301378. — 25. VALLOR S, VIERKANT T, MINDS MACH, 34 (2024) 20. doi:10.1007/s11023-024-09674-0. — 26. ACLU, Williams v. City of Detroit. https://www.aclu.org/cases/williams-v-city-of-detroit-face-recognition-false-arrest. — 27. DIRECTORATE-GENERAL FOR CITIZENS' RIGHTS, JUSTICE AND INSTITUTIONAL AFFAIRS (EUROPEAN PARLIAMENT), Artificial intelligence and civil liability: a European perspective, (Publications Office of the European Union 2025). — 28. SIO FS DE, MECACCI G, Philos Technol, 34 (2021) 1057. doi:10.1007/s13347-021-00450-x. — 29. RUDIN C, Nat Mach Intell, 1 (2019) 206. doi:10.1038/s42256-019-0048-x. — 30. FERGUSON A, Wash Univ Law Rev, 94 (2017) 1109. — 31. PARK S, NAN X, AI Soc, (2025). doi:10.1007/s00146-025-02620-3. — 32. NOUIS SC, UREN V, JARIWALA S, BMC Med Ethics, 26 (2025) 89. doi:10.1186/s12910-025-01243-z. — 33. CRAWFORD K, Atlas of AI (Yale University Press, 2021). doi:10.12987/9780300252392. — 34. EU, The Act Texts, EU Artificial Intelligence Act. https://artificialintelligenceact.eu/ai-act-explorer. — 35. PEREIRA M, GREENSTEIN S, SADUN R, TAMBE P, RONCHI DARRE L, GLAZER T, KIM A, DODHIA R, LAVISTA FERRES J, The New Digital Divide (National Bureau of Economic Research, 2024). doi:10.3386/w32932. — 36. DINKU T, Challenges with availability and quality of climate data in Africa. In: MELESSE AM, ABTEW W, SENAY G (Eds) Extreme Hydrology and Climate Variability (Elsevier, 2019). doi:10.1016/B978-0-12-815998-9.00007-5. — 37. CHINTA SV, WANG Z, PALIKHE A, ZHANG X, KASHIF A, SMITH MA, LIU J, ZHANG W, PLOS Digit Health, 4 (2025) e0000864. doi:10.1371/journal.pdig.0000864. — 38. TAO Y, VIBERG O, BAKER RS, KIZILCEC RF, PNAS Nexus, 3 (2024) 346. doi:10.1093/pnasnexus/pgae346. — 39. ALKHAMISSI B,

ELNOKRASHY M, ALKHAMISSI M, DIAB M, Investigating Cultural Alignment of Large Language Models. In: KU LW, MARTINS A, SRIKUMAR V (Eds) Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). (Association for Computational Linguistics, Bangkok, Thailand) 2024). doi:10.18653/v1/2024.acl-long.671. — 40. COULDRY N, MEJIAS UA, Telev New Media, 20 (2019) 336. doi:10.1177/1527476418796632. — 41. TAEIHAGH A, Policy Soc, 44 (2025) 1. doi:10.1093/polsoc/puaf001. — 42. TALLBERG J, ERMAN E, FURENDAL M, GEITH J, KLAMBERG M, LUNDGREN M, Int Stud Rev, 25 (2023) viad040. doi:10.1093/isr/viad040. — 43. CHO H, ACKOM E, Nat Commun, 16 (2025) 1228. doi:10.1038/s41467-024-53956-1. — 44. ELISH MC, Engag Sci Technol Soc Pre-Print, (2019). doi:10.17351/ests2019.260.

R. Dodhia

Microsoft AI for Good Lab, Seattle, USA
e-mail: rahuld@microsoft.com

UMJETNA INTELIGENCIJA ZA DRUŠTVENO DOBRO POČINJE S ODGOVORNOŠĆU

S A Ž E T A K

Umjetna inteligencija često se definira kroz pitanja koristi i rizika, no njezini učinci proizlaze iz načina na koji ona mijenja uvjete pod kojima se znanje proizvodi i na temelju kojih se djeluje. Ovaj rad ispituje umjetnu inteligenciju za društveno dobro kao skup praksi koje proširuju sposobnost promatranja, klasifikacije i zaključivanja na razinama izvan ljudske percepcije. Ti sustavi mijenjaju ono što postaje vidljivo, kako se ta vidljivost tumači i tko je pozicioniran da je koristi. Kako strojno zaključivanje preuzima veći dio promatračkog i interpretativnog tereta, autoritet se pomiče prema institucijama koje dizajniraju i upravljaju tim alatima. Analiza identificira odgovornost, transparentnost i jednakost kao praktične uvjete koji određuju hoće li umjetna inteligencija jačati ili slabiti zajednice i okruženja kojih se dotiče. Cilj je razjasniti kako umjetna inteligencija reorganizira odnos između promatranja, prosudbe i odgovornosti unutar projekata koji su definirani kao društveno korisni.