Chapter 7
'End-to-end' ethical artificial intelligence. Taking into account the social and natural environments of automation

Antonio A. Casilli

1. Introduction

In the last few decades, AI applications have largely been driven by machine learning paradigms. Instead of coding rules for every possible occurrence, machine learning algorithms can identify patterns in vast amounts of data to detect trends, find solutions and predict future events. Large language models like ChatGPT, recommendation algorithms on e-commerce platforms and operating systems for autonomous vehicles all use these techniques to learn from data.

Science, politics and industry have also focused on the ways in which these machine learning models influence behaviour. Although fears concerning artificial general intelligence are unrealistic, and an AGI (Artificial General Intelligence) that rivals or surpasses human cognition remains largely science fiction, AI's scale is especially concerning. Privacy and surveillance concerns are related to the data that these technologies require. Automation improves as information becomes richer, while data used for the development of machine learning models may include private or personally identifiable data. It is not only governments and police forces who conduct surveillance, but also businesses and the private sector.

The application of machine learning can, furthermore, pose a challenge to the transparency and democracy of decision-making since even the programmers cannot systematically determine which characteristics of the data the system has used in order to generate solutions. Machines can 'learn' from historical data that is biased to recognise, for instance, light-skinned men more accurately than dark-skinned women, to select men over women in recruitment processes or to predict reoffending risks that are higher for black than white defendants (Müller 2021).

Due to these recent advances, efforts to formulate ethical guidelines have mushroomed. In its crowdsourced AI Ethics Guidelines Global Inventory, the NGO AlgorithmWatch includes 173 such documents (AlgorithmWatch 2023). However, the most comprehensive analysis of 84 ethics charters has been published by three researchers at ETH Zurich, in Switzerland (Jobin et al. 2019). They identify five recurrent themes in their corpus: transparency; justice and fairness; non-maleficence; responsibility; and privacy. But each document defines these ethical domains differently and, more importantly, these principles tend to be operationalised mathematically and implemented as technical measures. Almost all the existing guidelines focus on machine learning's ability to resolve ethical issues (Hagendorff 2020). By reducing ethics principles to their technical

aspect, the proponents of ethical AI guidelines are neglecting broader socioeconomic, geographical and institutional factors.

The implications are profound and we must therefore take a holistic view, looking at the whole system in which AI is incorporated. The question to ask is: 'whose values are prioritised in these AI ethics guidelines?'. According to Anna Jobin and her colleagues, the geographical distribution of the issuers of ethics guidelines highlights hotspots and hubs in Europe and the United States, as well as Japan and India (China, which is another major player, was not included in the study). All these countries are heavily investing in artificial intelligence. On the contrary, countries from the Global South are absent from their map.

It is not a surprise, then, that AI ethics charters overlap geographically with producers of AI solutions in general, even though this may result in conflicts of interest. For example the principle of responsibility, if applied effectively, may interfere with the continuous exponential expansion of technologies and commercialisation. By the same token, minimising privacy risks may disrupt data collection and severely hamper machine learning. A strict adherence to ethical principles would clash with free market ideologies. A key responsibility of AI ethicists is to minimise such clashes.

Therefore, ethical AI is consistent with the tradition of science that addresses industry needs without challenging corporate motives. Existing guidelines are developed by or with the contribution of technology companies. As another means of serving the industrial status quo, AI ethicists refer to elite engineers as arbiters of 'bias' while excluding scholars and advocates who denounce power dynamics and economic imbalances. Meredith Whittaker suggests that corporate actors use ethical AI to 'co-opt and neutralise critique'. This is done in part by funding the 'weakest critics, often institutions and coalitions that focus on so-called AI ethics, and frame issues of tech power and dominance as abstract governance questions' (Whittaker 2021: 54).

Companies' implicit or explicit involvement in AI ethics fills the current regulatory vacuum – and ultimately contributes to it. The emphasis on in-house AI ethics prevents the development of legally mandated standards and enforcement mechanisms. The tech industry assumes it can formulate ethical norms for AI and ensure compliance, but the absence of any discussion of the effective ways of enforcing ethics standards in these charters proves that this assumption does not hold. Even when ethical principles are operationalised through specific tools, they fail to address existing imbalances. Another study conducted on 169 ethics documents finds that only 39 include AI ethics tools such as lists of best practice, checklists and adapted software applications. The most important aspect, however, is that key stakeholders were excluded from the design of these tools and that there was no external auditing (Ayling and Chapman 2022).

Admitting that AI ethics discourses are connected to the commercial interests of tech companies is the first step towards acknowledging that AI is industrially produced using human and natural resources. This industrial system requires appropriate corporate structures, capital investments and institutional backing. Within what context is AI produced? Who contributes and for what? When viewed from the standpoint of society

as a whole, what are its production costs? Virtually none of these questions are addressed in AI ethics charters which implicitly assume that voice assistants, recommendation engines and self-driving cars raise ethical concerns only when consumers use them. As in other areas of sociopolitical research on AI, ethics has historically put its emphasis on the possible effects of technology only at the deployment and in the marketing phases.

2. The AI production process

The production process behind AI is crucial long before the deployment of products and solutions. This is particularly clear if we consider the example of autonomous vehicles (Tubaro and Casilli 2019). Besides engineers, software developers and designers, safety drivers are also needed. These drivers travel inside the car, monitor the trip, provide feedback to the technical team and are expected to take control of the vehicle if necessary. The development of self-driving cars, however, also requires a vast army of hidden workers. Some refer to these as 'data workers' since they manage information; others as 'microworkers' since their tasks are fragmented and are viewed as less important than those of data scientists and software developers. For autonomous vehicles to be safe, computer-vision algorithms must be capable of recognising pedestrians crossing the street, for example. How does the car know what a pedestrian looks like? Generally, these examples are based on large sets of image data. The images routinely taken by cameras and sensors mounted on autonomous cars provide precisely this. However, these images need to be labelled before they can be used. In a traffic photo, the computer needs to read tags indicating 'pedestrian', 'bike', 'traffic light', 'bus', etc. Human workers are paid to add these tags, thus making everything visible to the AI. It is a huge job because the car's algorithm cannot learn from small amounts of data. Annotation would be tedious and lengthy if only a few workers were involved. But by fragmenting these large batches into many short, one-shot tasks, and assigning them to many data workers, each of whom perform just one or two, the goal can be achieved.

Human work is performed off-street by data annotators who remotely tag the images for the development of autonomous cars. Under what conditions are these human workers recruited, remunerated, managed and facilitated in exercising their rights?

Autonomous cars do not represent an isolated case when it comes to production-related ethics issues, but rather exemplify the trends seen throughout AI. Human labour and the environment are the two main issues that arise.

Despite the high value the AI industry places on its visible workforce of software developers, data scientists and computer engineers, it tends to ignore and to render invisible its lower-level microworkers who are indispensable, despite performing repetitive and often unqualified data tasks. It is these invisible humans that intervene at various stages during the development process of machine learning models: the initial training of the models (data generation and annotation); the verification of their

outputs after deployment; and, sometimes, performing the real-time correction or 'impersonation' of AI systems.[1]

Mechanical Turk, Amazon's microtasking platform, popularised human-powered data work for AI in the middle of the first decade of the 2000s. The name comes from an Ottoman-dressed chess-playing automaton from the eighteenth century. This 'proto-AI' could supposedly simulate the cognitive processes of a real chess player. The original mechanical Turk, however, was a hoax, controlled by a hidden operator. Today, it serves as a metaphor for the 'human-in-the-loop' principle that governs AI production. Human workers still prepare, test and sometimes pose as autonomous systems, but on a much larger scale. Amazon Mechanical Turk has hundreds of thousands of freelance workers who perform human intelligence tasks (HITs) which are fairly straightforward for humans but difficult for machines: recognising objects, creating lists, transcribing short sentences, etc. This microwork has been described by Amazon's founder Jeff Bezos, without a hint of irony, as 'artificial artificial intelligence' (Casilli and Posada 2019).

There have been many up and coming companies entering this market since Mechanical Turk launched. In addition to the Australian Appen, the American Remotasks or the German Clickworker, several international platforms provide microworking services on demand. Technology multinationals have created their own microworking services where they act as the sole recruiter, such as Microsoft's UHRS (Universal Human Relevance System) and Google's RaterHub. Companies that operate as BPO (business process outsourcing) vendors can also recruit workers for their clients in countries where the workforce is cheaper, such as India or Venezuela. Among them are some very big companies and platforms; others are smaller and sometimes specialised, for instance IsAHit and Wirk in France. A few smaller start-ups have been acquired by larger companies, such as Mighty AI, which is dedicated to the automotive industry and was acquired by Uber Advanced Technologies Group in 2019. Subsequently, the entire ATG was acquired by a startup, Aurora, in which Uber holds a 40% ownership stake. Companies, platforms and startups are all involved in complex arrangements which shows that tech companies need to be agile. A large number of intermediaries must be used by AI developers to recruit disposable workers.

3. Invisibility and the working conditions of microworkers

The conditions under which this type of work is performed are problematic.

Microworkers rarely figure on the payroll: typically, platforms bind them through membership or participation contracts similar to the general terms of service that are routinely agreed to by consumers of internet services and mobile apps. This leaves workers vulnerable to market volatility and without social protection. In addition, they work remotely from anywhere, which opens them up to worldwide competition and lower wages. The common practice of paying by the piece rather than by the hour or by a monthly salary makes it difficult to control, even though some platforms (such as

1. This three-part conceptual framework is further developed in Tubaro et al. (2020).

Clickworker) recommend paying the minimum wage. Microtasks can be paid as little as a few cents in the most dire cases. In a report published in 2018, the International Labour Organization estimated that, on average, microworkers earn 3.31 dollars per hour (counting the time they spend searching for new tasks to perform), well below the minimum wage in most countries (Berg et al. 2018). TIME revealed in January 2023 that workers on one of OpenAI's subcontracting platforms in Kenya were earning only between 1.34 and 2 dollars per hour when they were annotating data for ChatGPT (Perrigo 2023).

Economic necessity often motivates the workers who perform these data tasks in spite of the low wages. The situation is not limited to low- and middle-income countries. Other research has found that low wages and poverty-level workers are over-represented even among microworkers in France, a high income country (Casilli et al. 2019). Women with young children often work part-time and supplement their income with microtasks. There are no clear indications that they can derive additional benefits in terms of career progression; in the future their skills won't make them particularly attractive to employers (Tubaro et al. 2022).

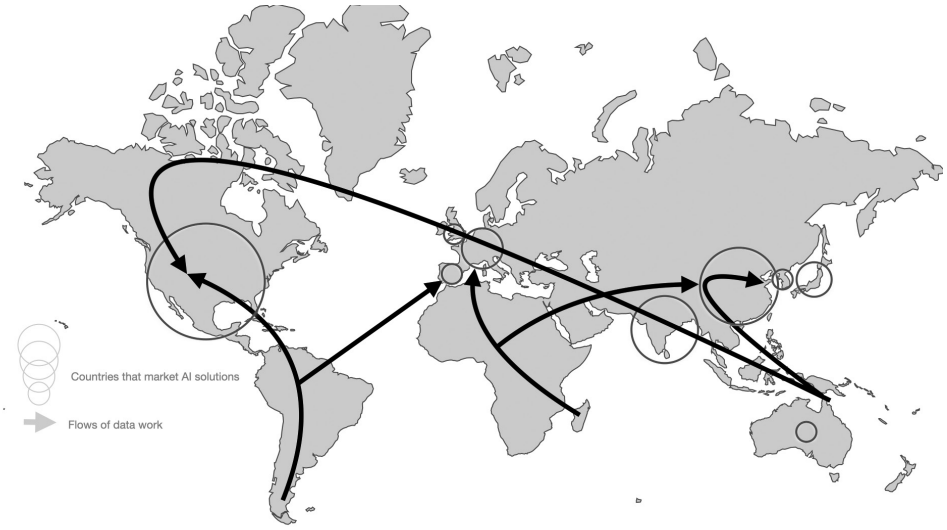
This contradicts some of the fundamental principles outlined in the AI ethics charters discussed above, including fairness, justice and transparency – and even privacy, as workers are sometimes asked to provide personal data, such as selfies and voice recordings, for the creation of machine learning datasets. The industry cannot even meet its own standards, regardless of how vague they may be in the first place.

Microwork's invisibility, low wages and precarious status are problematic, as its geographical distribution makes all the more clear. In contrast to charters and guidelines, which are usually published in high income countries, ongoing research on data production shows what a flipped image of ethical AI looks like. The majority of microworkers are located in the Global South, despite it being Global North countries that have attracted the most research attention (Difallah et al. 2018). Efforts to map their global distribution suggest that they reproduce legacy inequalities based on wealth, power and geographical influence (Graham et al. 2017), although this is perhaps a natural conclusion for results that were obtained by analysing English-speaking platforms, where clear links appear with former British colonies, such as India, or former zones of US hegemony, such as the Philippines (Gray and Suri 2019).

Based on these pioneering studies, team members at DiPLab (Digital Platform Labour at the Institut Polytechnique de Paris) have conducted extensive research in French-speaking African countries (including Madagascar, Cameroon, Mali, Senegal, Morocco and Egypt), as well as Portuguese and Spanish-speaking Latin American countries which mainly serve North American tech companies (Le Ludec et al. 2023; Viana Braz et al. 2023).[2] Venezuela, Argentina (Miceli and Posada 2022) and Brazil (Grohman and Fernandes Araújo 2021) are among the most active countries in this international market.

2. Results were obtained within the framework of DiPLab's projects HUSH (Human Supply Chain of Smart Technologies), funded by the French National Research Agency (ANR); and TRIA (El Trabajo de la Inteligencia Artificial), funded by the French Centre for Scientific Research (CNRS).

Figure 1 **Global flows of annotated data from microwork providers to AI solution providers**



Source: DiPLab's HUSH and TRIA projects. Author's elaboration.

On the basis of current evidence, Figure 1 schematically represents the global flows of data and work that are feeding the development of AI. The Latin American workforce serves technology producers in North America (particularly the United States) and Europe. From South and Southeast Asia, work primarily goes to North America but also to China. Africa provides data work to Europe and China. The latter has substantial internal flows, despite little knowledge of them, as do both Europe (with some east-west flows) and the United States. As the exact size of the flows is still not documented in many cases, the map is only indicative. However, it reveals a major flaw in current AI ethics approaches: the severe underrepresentation of the Global South in the creation of charters and guidelines, as well as its overrepresentation within the 'human supply chain' that produces AI and undermines diversity and cultural awareness – and, consequently, reduces the voice of workers.

4. Natural resources

Along with the labour force required to annotate and enrich data, AI consumes natural resources such as energy, minerals and metals in its construction as well as energy in the performance of heavy calculations. This raises questions about sustainability, a topic rarely mentioned in the charters and guidelines.

For AI to be effective, it must respect the natural environment where resources are used as well as the human environment where labour is produced – which explains why recent years have seen a rise in attention to the environmental costs of AI. The challenge has been met in two fundamentally different ways: by quantifying the carbon footprint

of computational processes; and by denouncing the extractivist logic underlying the tech industry.

Several efforts have been made to measure the environmental cost of machine learning. Training one large Natural Language Processing (NLP) transformer model generates nearly five times the amount of carbon dioxide of a single car's annual emissions, or 50 times the amount emitted by one individual in a lifetime (Strubell et al. 2019). Various trackers and impact measurements help researchers assess the energy use of their tools and make actionable recommendations to reduce emissions (Bannour et al. 2021). Many companies, including Alphabet, are investing heavily in green energy sources and developing more efficient ways to cool their data centres. AI training can also be enhanced with 'green algorithms', efficient architectural settings and smaller models (Cai et al. 2019). Even though these efforts are commendable, it must be acknowledged that, once again, they rely on self-regulation and assume that the tech industry can and will follow. Therefore, one must assume that the technological systems that pollute and waste natural resources also have the potential to mitigate them.

Yet, over-reliance on portmanteau concepts such as 'environment', 'climate' and 'energy' obscures the concrete human and material substratum that is supporting the transformation of natural resources into AI. Other researchers have thus developed a more radical theoretical perspective centred on digital extractivism to analyse the materiality of AI, its reliance on natural resources and its links to capitalism (see Brevini in this volume; Iyer et al. 2021). Rather than referring to extraction as a plundering of natural resources, they use the notion to describe how capital interacts with and draws on human, political, economic and social activity. In contemporary capitalism, extractivism is ubiquitous, spanning not only traditional sectors like logistics and agriculture but also intangible activities like finance.

Data production and AI are new frontiers of extraction in this sense. Algorithms and data would not function without the minerals and metals that form the core hardware components that host and compute them. Several countries, including Zimbabwe, Madagascar, Bolivia, Argentina and Chile, mine minerals like cobalt, nickel and lithium for batteries. Indonesia is a source of tin for high income regions and countries like Europe and the US, while those that produce semiconductors, like Taiwan and South Korea, can be seen in the same light. Author Kate Crawford calls the 'mineralogical layer of AI' the foundation of the informational infrastructure that fuels intelligent solutions (Crawford 2021). The environmental costs of this must also be taken into account in the moving of minerals, metals, fuel, hardware and final products internationally.

A map of the locations where most natural resources are extracted for the AI industry overlaps significantly with the trajectories of microwork flows. As data workers from the Global South are mobilised for the AI industry of the North, those who work in mining, transport and electronic waste management follow the same global patterns. The production of information technologies is also part of an international division of labour (Fuchs 2016). Information and contents circulating from remote servers to our screens are not just produced by data processing activities. Instead, they revolve around the extraction and transportation of the minerals used in electronics.

Today's production processes for AI have significant limitations that would go unnoticed if we only focused on the principles emphasised in current public debates and ethics guidelines. But there are ways of overcoming these deficiencies. In accordance with the analyses advanced by Kate Crawford (2021), Gunay Kazimzade and Milagros Miceli (2020), a new, alternative research programme would adopt a dual approach to AI by taking into account not just its technological but its social, economic and political context as well. In addition to the now popular depiction of AI as a data intensive technology, it raises the important question of how such data is created.

5. Conclusion

End-to-end ethical AI requires a consideration of the conditions of production of the data, tools and equipment used to manufacture and market these systems. A number of consumer products already apply this kind of ethical reasoning. For example, footwear and processed food manufacturers do not discriminate against their consumers based on gender, location or other factors – they are not as biased as AI in this respect. However, that does not make them ethical. A company is considered ethical if it respects workers' rights, provides decent working conditions and minimises its environmental impact. In a similar vein, an ethical AI must minimise negative externalities both within and outside human communities and the same with regard to its production processes.

Taking account of the human, social, political and economic contexts of today's AI technologies can provide novel insights and suggest directions for future action. It is important to protect the rights of remote AI workers. They should be able to resist unsuitable conditions if they have the opportunity. They should be able to protest against technological systems they consider ethically problematic and they should be able to refuse to contribute to them. In such a scenario, AI ethics shifts from promoting the interests of producers-owners to promoting the ethical agency of producers-workers. For this to succeed, it is necessary to acknowledge the invisible work of preparation, verification and impersonation of automated systems in order to establish a stronger, more fundamental approach to AI ethics. As a result, workers should be provided with methods of protection. Other workers directly and indirectly involved in the supply chain of modern computing devices could benefit from the same principles.

References

AlgorithmWatch (2023) AI ethics guidelines global inventory. https://inventory.algorithmwatch.org/

Ayling J. and Chapman A. (2022) Putting AI ethics to work: Are the tools fit for purpose?, AI and Ethics, 2, 405–429. https://doi.org/10.1007/s43681-021-00084-x

Bannour N., Ghannay S., Névéol A. and Ligozat A.L. (2021) Evaluating the carbon footprint of NLP methods: A survey and analysis of existing tools, Proceedings of the Second Workshop on Simple and Efficient Natural Language Processing, 11–21. https://aclanthology.org/2021.sustainlp-1.2/

Berg J., Furrer M., Harmon E., Rani U. and Silberman M.S. (2018) Digital labour platforms and the future of work: Towards decent work in the online world, ILO. https://www.ilo.org/global/publications/books/WCMS_645337/lang--en/index.htm

Cai H., Gan C., Wang T., Zhang Z. and Han S. (2019) Once-for-all: Train one network and specialize it for efficient deployment. https://doi.org/10.48550/arXiv.1908.09791

Casilli A.A. and Posada J. (2019) The platformization of labor and society, in Graham M. and Dutton W.H. (eds.) (2019) Society and the Internet. How networks of information and communication are changing our lives, 2nd ed., Oxford University Press, 293–306. https://doi.org/10.1093/oso/9780198843498.003.0018

Casilli A.A., Tubaro P., Le Ludec C., Coville M., Besenval M., Mouhtare T. and Wahal E. (2019) Le micro-travail en France. Derrière l'automatisation, de nouvelles précarités au travail ?, Projet de recherche DiPLab.

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

Difallah D., Filatova E. and Ipeirotis P. (2018) Demographics and dynamics of mechanical Turk workers, Proceedings of the eleventh ACM international conference on web search and data mining, 135–143. https://doi.org/10.1145/3159652.3159661

Fuchs C. (2016) Digital labor and imperialism, Monthly Review, 67 (8), 14–24. https://doi.org/10.14452/MR-067-08-2016-01_2

Graham M., Hjorth I. and Lehdonvirta V. (2017) Digital labour and development: Impacts of global digital labour platforms and the gig economy on worker livelihoods, Transfer, 23 (2), 135–162. https://doi.org/10.1177/1024258916687250

Gray M.L. and Suri S. (2019) Ghost work: How to stop Silicon Valley from building a new global underclass, Houghton Mifflin Harcourt.

Grohmann R. and Fernandes Araújo W. (2021) Beyond mechanical Turk: The work of Brazilians on global AI platforms, in Verdegem P. (ed.) AI for everyone? Critical perspectives, University of Westminster Press, 247–266. https://library.oapen.org/bitstream/handle/20.500.12657/58191/1/ai-for-everyone.pdf#page=256

Hagendorff T. (2020) The ethics of AI ethics: An evaluation of guidelines, Minds and Machines, 30, 99–120. https://doi.org/10.1007/s11023-020-09517-8

Iyer N., Achieng G., Borokini F. and Ludger U. (2021) Automated imperialism, expansionist dreams: Exploring digital extractivism in Africa, Pollicy. https://archive.pollicy.org/wp-content/uploads/2021/06/Automated-Imperialism-Expansionist-Dreams-Exploring-Digital-Extractivism-in-Africa.pdf

Jobin A., Ienca M. and Vayena E. (2019) The global landscape of AI ethics guidelines, Nature Machine Intelligence, 1, 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kazimzade G. and Miceli M. (2020) Biased priorities, biased outcomes: Three recommendations for ethics-oriented data annotation practices, Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 71. https://doi.org/10.1145/3375627.3375809

Le Ludec C., Cornet M. and Casilli A.A. (2023) The problem with annotation. Human labour and outsourcing between France and Madagascar, Big Data and Society, 10 (2). https://doi.org/10.1177/20539517231188723

Miceli M. and Posada J. (2022) The data-production dispositif, Proceedings of the ACM on Human-Computer Interaction, 6 (CSCW2), 1–37. https://doi.org/10.1145/3555561

Müller V.C. (2021) Ethics of artificial intelligence and robotics, in Zalta E.N. and Nodelman U. (eds.) The Stanford encyclopedia of philosophy, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/

Perrigo B. (2023) Exclusive: OpenAI used Kenyan workers on less than $2 per hour, Time, 18 January 2023. https://time.com/6247678/openai-chatgpt-kenya-workers/
Strubell E., Ganesh A. and McCallum A. (2019) Energy and policy considerations for deep learning in NLP. https://doi.org/10.48550/arXiv.1906.02243
Tubaro P. and Casilli A.A. (2019) Micro-work, artificial intelligence and the automotive industry, Journal of Industrial and Business Economics, 46 (3), 333–345. https://doi.org/10.1007/s40812-019-00121-1
Tubaro P., Casilli A.A. and Coville M. (2020) The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence, Big Data and Society, 7 (1). https://doi.org/10.1177/2053951720919776
Tubaro P., Coville M., Le Ludec C. and Casilli A.A. (2022) Hidden inequalities: The gendered labour of women on micro-tasking platforms, Internet Policy Review, 11 (1), 1–26. https://doi.org/10.14763/2022.1.1623
Viana Braz M., Tubaro P. and Casilli A.A. (2023) Microwork in Brazil: Who are the workers behind AI?, Research Report DiPLab and LATRAPS. https://hal.science/hal-04140411
Whittaker M. (2021) The steep cost of capture, Interactions, 28 (6), 50–55. https://doi.org/10.1145/3488666

All links were checked on 29.01.2024.

Cite this chapter: Casilli A.A. (2024) ‘End-to-end’ ethical artificial intelligence. Taking into account the social and natural environments of automation, in Ponce del Castillo A. (ed.) Artificial intelligence, labour and society, ETUI.