



## Le regulatory sandboxes come strumento etico per un'intelligenza artificiale affidabile\*

Corrado Claverini

Università degli Studi del Salento

L'AI Act rappresenta un passo significativo verso la regolamentazione dell'intelligenza artificiale all'interno dell'Unione Europea, introducendo categorie di rischio, da “minimo” a “inaccettabile”, per le applicazioni di tale tecnologia. Pubblicato sulla *Gazzetta Ufficiale dell'Unione Europea* il 12 luglio 2024,<sup>1</sup> l'AI Act è entrato in vigore il 1° agosto 2024. Non è l'unica legge sull'intelligenza artificiale esistente a livello globale e forse – come è stato notato<sup>2</sup> – non è neanche la prima, ma è senz'altro uno dei tentativi più organici che siano mai stati effettuati a garanzia di una innovazione responsabile in tale ambito. Negli Stati Uniti, ad esempio, si è provveduto con un ordine esecutivo (30 ottobre 2023)<sup>3</sup> e un memorandum (24 ottobre 2024)<sup>4</sup> del Presidente Biden. Prima ancora, in Cina, sono entrate in vigore le *Misure provvisorie per la gestione dei servizi di intelligenza artificiale generativa* (15 agosto 2023).<sup>5</sup> Ma, al di là delle primogeniture, l'AI Act – con i suoi 113 articoli, suddivisi in 13 capi, e 13 allegati – ha alimentato, in maniera forte, il dibattito sul trade off fra innovazione e regolamentazione.<sup>6</sup> Sono varie le domande che sono state sollevate a tal riguardo: questa legge europea frena in qualche modo l'innovazione? Oppure, al contrario, la promuove in maniera responsabile? Si tratta di interrogativi cui non è facile rispondere univocamente, ma – come è stato rilevato<sup>7</sup> – una delle misure più interessanti a

\* Questo articolo fa parte del progetto AInCP che ha ricevuto finanziamenti dalla Commissione Europea, Horizon Europe Research and Innovation Action, nell'ambito del Grant Agreement n. 101057309.

<sup>1</sup> Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale) <<https://eur-lex.europa.eu/eli/reg/2024/1689>> (d'ora in poi: AI Act).

<sup>2</sup> V. Franceschelli, “Norme strutturali e organizzative ad alto impatto sui diversi settori”, *Dossier 3 (2024) – Intelligenza artificiale e il nuovo AI Act europeo*, 12 settembre 2024 <<https://ntplusdiritto.ilsole24ore.com/art/le-novita--norme-strutturali-e-organizzative-ad-alto-impatto-diversi-settori-AF2UvXqD>>.

<sup>3</sup> *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* <<https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>>.

<sup>4</sup> *Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence* <<https://www.whitehouse.gov/briefing-room/presidential-actions/2024/10/24/memorandum-on-advancing-the-united-states-leadership-in-artificial-intelligence-harnessing-artificial-intelligence-to-fulfill-national-security-objectives-and-fostering-the-safety-security/>>.

<sup>5</sup> *Interim Measures for the Management of Generative AI Services* <[https://www.cac.gov.cn/2023-07/13/c\\_1690898327029107.htm](https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm)>.

<sup>6</sup> K.J.D. Chan, G. Papsyshev, M. Yarime, “Balancing the Tradeoff between Regulation and Innovation for Artificial Intelligence: An Analysis of Top-down Command and Control and Bottom-up Self-Regulatory Approaches”, *SSRN* (20 ottobre 2022) <<http://dx.doi.org/10.2139/ssrn.4223016>>.

<sup>7</sup> M. Tupponi, “Le misure a sostegno dell'innovazione – Sperimentazione e sandbox: le norme a favore delle imprese”, *Dossier 4 (2024) – Intelligenza artificiale e il nuovo AI Act europeo*, 4 novembre 2024 <<https://ntplusdiritto.ilsole24ore.com/art/le-misure-sostegno-innovazione--sperimentazione-e-sandbox-norme-favore-imprese-AGQU9au>>.



sostegno dell'innovazione previste da tale regolamento è l'introduzione delle regulatory sandboxes o spazi di sperimentazione normativa.<sup>8</sup> Secondo l'articolo 57 dell'AI Act, le sandboxes normative

garantiscono un ambiente controllato che promuove l'innovazione e facilita lo sviluppo, l'addestramento, la sperimentazione e la convalida di sistemi di IA innovativi per un periodo di tempo limitato prima della loro immissione sul mercato o della loro messa in servizio conformemente a un piano specifico dello spazio di sperimentazione concordato tra i fornitori o i potenziali fornitori e l'autorità competente. Tali spazi di sperimentazione possono comprendere prove in condizioni reali soggette a controllo nei medesimi spazi.<sup>9</sup>

Insomma, lungi dal limitare l'innovazione, l'AI Act la promuove in modo responsabile. Non a caso, il Capo VI dell'AI Act è intitolato *Misure a sostegno dell'innovazione* (artt. 57-63) e prevede l'introduzione di questi ambienti controllati in cui le imprese possono testare sistemi di intelligenza artificiale prima della loro immissione sul mercato e mitigare gli eventuali rischi. Qualora l'attenuazione di tali rischi non fosse possibile, «le autorità nazionali competenti hanno il potere di sospendere, in via temporanea o permanente, il processo di prova o la partecipazione allo spazio di sperimentazione».<sup>10</sup> Inoltre, è importante rilevare come le imprese siano sì «responsabili per eventuali danni arrecati a terzi a seguito della sperimentazione che ha avuto luogo nello spazio di sperimentazione», ma – se avranno rispettato «il piano specifico e i termini e le condizioni di partecipazione», nonché «gli orientamenti forniti dall'autorità nazionale competente»<sup>11</sup> – non saranno soggette a sanzioni amministrative pecuniarie. Infine, va notato come «l'accesso agli spazi di sperimentazione normativa per l'IA sia gratuito per le PMI, comprese le start-up».<sup>12</sup>

L'idea è dunque quella di valutare, fin dalle prime fasi, i possibili rischi legati allo sviluppo di una tecnologia che faccia uso di IA e abbracciare un concetto di innovazione responsabile in linea con lo spirito che ha animato il percorso che ha condotto all'approvazione dell'AI Act e che – secondo quanto stabilito dall'articolo 113 di tale regolamento – porterà le autorità competenti degli Stati membri a istituire, entro il 2 agosto 2026, almeno una regulatory sandbox in materia di intelligenza artificiale a livello nazionale.

L'Italia si sta già muovendo in tale direzione. Il Consiglio dei ministri n. 78 del 23 aprile 2024 ha approvato, infatti, il disegno di legge *Disposizioni e delega al Governo in materia di intelligenza artificiale*.<sup>13</sup> Presentato in Senato il 20 maggio 2024, tale disegno di legge mira a promuovere «un

<sup>8</sup> L'articolo 3(55) dell'AI Act definisce uno «spazio di sperimentazione normativa per l'IA» in questo modo: «un quadro controllato istituito da un'autorità competente che offre ai fornitori o potenziali fornitori di sistemi di IA la possibilità di sviluppare, addestrare, convalidare e provare, se del caso in condizioni reali, un sistema di IA innovativo, conformemente a un *piano dello spazio di sperimentazione* per un periodo di tempo limitato sotto supervisione regolamentare» (corsivo mio). A sua volta un «piano dello spazio di sperimentazione» è «un documento concordato tra il fornitore partecipante e l'autorità competente in cui sono descritti gli obiettivi, le condizioni, il calendario, la metodologia e i requisiti relativamente alle attività svolte all'interno dello spazio di sperimentazione» (cfr. articolo 3(54) dell'AI Act). Infine, un «piano di prova in condizioni reali» è «un documento che descrive gli obiettivi, la metodologia, l'ambito geografico, della popolazione e temporale, il monitoraggio, l'organizzazione e lo svolgimento della prova in condizioni reali» (cfr. articolo 3(53) dell'AI Act).

<sup>9</sup> AI Act, articolo 57(5).

<sup>10</sup> AI Act, articolo 57(11).

<sup>11</sup> AI Act, articolo 57(12).

<sup>12</sup> AI Act, articolo 58(2)d.

<sup>13</sup> Disposizioni e delega al Governo in materia di intelligenza artificiale – Atto Senato n. 1146 <[https://www.senato.it/leg/19/BGT/Schede/Ddliter/testi/58262\\_testi.htm](https://www.senato.it/leg/19/BGT/Schede/Ddliter/testi/58262_testi.htm)> (d'ora in poi DDL 1146). Su questo si veda A. Sirotti Gaudenzi, «La ricerca di regole interne – Il 'Bel Paese' rincorre l'Europa nel mettere a punto una disciplina», *Dossier*



utilizzo corretto, trasparente e responsabile, in una dimensione antropocentrica, dell'intelligenza artificiale, volto a coglierne le opportunità». <sup>14</sup> Secondo quanto stabilito dall'articolo 18, «l'Agenzia per l'Italia digitale (AgID) e l'Agenzia per la cybersicurezza nazionale (ACN) sono designate quali Autorità nazionali per l'intelligenza artificiale». <sup>15</sup> Tali Autorità,

ciascuna per quanto di rispettiva competenza, assicurano l'istituzione e la gestione congiunta di *spazi di sperimentazione* finalizzati alla realizzazione di sistemi di intelligenza artificiale conformi alla normativa nazionale e dell'Unione europea, sentito il Ministero della di fesa per gli aspetti relativi ai sistemi di intelligenza artificiale impiegabili in chiave duale. <sup>16</sup>

Anche l'Italia, dunque, avrà presto almeno una regulatory sandbox in materia di intelligenza artificiale. Ma, in altri settori, tale strumento non è una novità, essendo nato nel Regno Unito nel 2015 e, in seguito, diffusosi rapidamente in tutto il mondo. <sup>17</sup> In Italia è stato introdotto con il Decreto ministeriale 30 aprile 2021, n. 100, per la sperimentazione in ambito fintech, ovvero «le attività volte al perseguimento, mediante nuove tecnologie, dell'innovazione di servizi e di prodotti nei settori bancario, finanziario, assicurativo». <sup>18</sup> Gli operatori del settore possono fare domanda per l'accesso alla sperimentazione alle autorità di vigilanza: Consob, IVASS e Banca d'Italia. <sup>19</sup>

Nato, come si diceva, nel Regno Unito nel 2015 in seguito alle raccomandazioni della Financial Conduct Authority (FCA), <sup>20</sup> questo strumento è stato valutato positivamente sin da subito. <sup>21</sup> Ma, in seguito alla diffusione globale dell'istituto della regulatory sandbox, il Regno Unito ha incaricato Ron

---

4 (2024) – *Intelligenza artificiale e il nuovo AI Act europeo*, 4 novembre 2024 <<https://ntplusdiritto.ilsole24ore.com/art/la-ricerca-regole-interne--bel-paese-rincorre-europa-mettere-punto-disciplina-AG62vYu>>.

<sup>14</sup> DDL 1146, articolo 1(1).

<sup>15</sup> DDL 1146, articolo 18(1).

<sup>16</sup> DDL 1146, articolo 18(1)c, corsivo mio.

<sup>17</sup> Su questo si veda, in particolare, A. Merlino, *Regulatory Sandbox. Una nuova prospettiva ordinamentale*, Editoriale Scientifica, Napoli, 2022.

<sup>18</sup> Regolamento recante attuazione dell'articolo 36, commi 2-bis e seguenti, del decreto-legge 30 aprile 2019, n. 34, convertito, con modificazioni, dalla legge 28 giugno 2019, n. 58, sulla disciplina del Comitato e della sperimentazione FinTech (*Gazzetta Ufficiale Serie Generale* 157 (2 luglio 2021)) <<https://www.gazzettaufficiale.it/eli/id/2021/07/02/21G00109/sg>>.

<sup>19</sup> L'iter che ha portato all'emanazione del decreto non è stato senza dibattito. Si veda in particolare la preoccupazione della Consob su soggetti la cui sede non sia in Italia: «La CONSOB, con nota prot. n. 746661/20 del 30 luglio 2020, ha invece espresso parere negativo limitatamente alla previsione che amplia l'ambito di applicazione della sperimentazione a soggetti che non abbiano la loro sede in Italia, e ciò in ragione della riduzione delle competenze e dei poteri esercitabili e della conseguente ridotta portata dell'attività di vigilanza esercitabili da parte dell'Autorità dello Stato ospitante. In proposito, il Ministero ha rappresentato di aver, tuttavia, ritenuto opportuno non escludere dall'applicazione del regolamento in oggetto le attività svolte da o a favore di soggetti con sede legale in uno Stato dell'U.E. e operanti in Italia in regime di libera prestazione di servizi»

<[https://portali.giustizia-amministrativa.it/portale/pages/istituzionale/visualizza/?nodeRef=&schema=consul&nrg=202001523&nomeFile=202100109\\_27.html&subDir=Provvedimenti](https://portali.giustizia-amministrativa.it/portale/pages/istituzionale/visualizza/?nodeRef=&schema=consul&nrg=202001523&nomeFile=202100109_27.html&subDir=Provvedimenti)>.

<sup>20</sup> «We recommend considering these options further when we have collected more evidence about the specific needs of innovative firms during the testing of new solutions. This evidence would enable us to understand better which specific rules or legislation firms struggle with when testing innovative solutions and then analyse the effect the above legislative changes could have» <<https://www.fca.org.uk/publication/research/regulatory-sandbox.pdf>>.

<sup>21</sup> Nel *Regulatory sandbox lessons learned report* (2017) della FCA si legge: «the sandbox has allowed us to work with innovators to build appropriate consumer protection safeguards into new products and services» <<https://www.fca.org.uk/publication/research-and-data/regulatory-sandbox-lessons-learned-report.pdf>>.



Kalifa di elaborare un piano strategico per mantenere la leadership in questo settore.<sup>22</sup> Oltre al suggerimento di passare dalla sandbox alla scalebox,<sup>23</sup> la cosiddetta *Kalifa review* invitava a rendere questo strumento permanente,<sup>24</sup> nonché a supportare anche proposte di innovazione soltanto parziale.<sup>25</sup>

Oltre all'ambito fintech, le regulatory sandboxes sono state utilizzate ampiamente anche in campo sanitario e, in particolare, nei processi di Health Technology Assessment (HTA).<sup>26</sup> Ma, al di là dei numerosi vantaggi che offre questo strumento in termini di innovazione tecnologica, sono presenti delle criticità che non sono passate inosservate. Si pensi agli spazi di sperimentazione normativa per l'IA previsti dall'AI Act. Se da una parte le autorità di vigilanza potranno fornire un servizio di consulenza ai soggetti partecipanti alla sperimentazione in merito alla conformità dei loro sistemi di intelligenza artificiale al nuovo regolamento europeo,<sup>27</sup> dall'altra c'è il rischio concreto di dare priorità

---

<sup>22</sup> «The FCA's regulatory sandbox has played an instrumental role in supporting innovation and encouraging UK fintech business. The UK is a welcoming business and regulatory environment for fintechs. However, as other jurisdictions are launching their own initiatives, the UK needs to take steps to ensure that its own offering remains market-leading» <<https://assets.publishing.service.gov.uk/media/607979c7d3bf7f400f5b3c65/KalifaReviewofUKFintech01.pdf>>.

<sup>23</sup> «Support fintech businesses and innovation through a “scalebox” arrangement under which firms focussing on innovative technology receive additional support (particularly in their growth phase and/or in relation to the Priority Fintech Areas)» (*ibidem*).

<sup>24</sup> «We recommend that a permanent “digital sandbox” should be created. The digital sandbox should: a. focus on the Priority Fintech Areas (but without excluding the possibility of supporting other areas); b. allow digital collaboration, access to synthetic data sets, design and deployment on open source and open architecture “plug and play” at international, national and sectoral level; c. in the short to medium term, be led by the FCA. Through its TechSprint programme and related initiatives, the FCA has extensive experience in convening the financial services industry to help solve industry-wide problems and is best-placed to further develop the design and nature of a permanent digital sandbox offering. However, there are also additional approvals that are often required within organisations to participate in projects lead by the regulator. In the medium to long term the initiative is likely to benefit from being housed within a neutral, non-regulatory body, whilst still involving the active participation of the FCA. The FCA would offer guidance, co-ordinate with other stakeholders and assist with the creation of synthetic data sets and realistic inputs; d. allow access not only for fintechs, but for established regulated firms who have innovative technologies or products that they wish to test; e. provide access to a broader range of tools, including professional and other business services; f. have a high degree of transparency, where appropriate – with the results of its work potentially being available to the public or to approved observers; and g. involve the provision of physical spaces or “hubs” for this collaboration to take place. It is important that the digital sandbox should be available to participants across the UK and geographically-spread hubs could make the digital sandbox more widely accessible» (*ibidem*).

<sup>25</sup> «The regulatory sandbox should be enhanced in the following ways: a. by being available on a rolling basis, rather than through timelimited windows; b. by offering support even where the proposal is not the first of a kind but is still providing and/or delivering an innovative proposition to the market; and c. by creating a dedicated space in the regulatory sandbox for the Priority Fintech Areas, with specialist support where necessary. The FCA's regulatory sandbox is already the busiest sandbox facility in the world, and these proposals may increase demand for it. If necessary, some reconsideration of the FCA's objectives may also be necessary to allow the FCA to operate in the manner described above» (*ibidem*).

<sup>26</sup> Per l'ambito sanitario si veda E. Leckenby, D. Dawoud, J. Bouvy, P. Jonsson, “The Sandbox Approach and its Potential for Use in Health Technology Assessment: A Literature Review”, *Applied Health Economics and Health Policy* 19, 6, (2021), pp. 857-869, doi: 10.1007/s40258-021-00665-1. Cfr. anche F. Battaglia, G. Di Vetta, *Technology to unlock the mind: citizen science and the sandbox approach for a new model of BCI governance*, in 2022 *IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE)*, pp. 563-567, doi: 10.1109/MetroXRINE54828.2022.9967580.

<sup>27</sup> Si veda D. Baldini, K. Francis, “AI Regulatory Sandboxes between the AI Act and the GDPR: the role of Data Protection as a Corporate Social Responsibility”, *CEUR Workshop Proceedings* 3731 (2024) <<https://cris.maastrichtuniversity.nl/ws/portalfiles/portal/213129909/paper07.pdf>>.



all'innovazione piuttosto che mettere in atto misure di salvaguardia adeguate a tutelare i consumatori.<sup>28</sup> Dunque, è vero che coloro che avranno accesso a una regulatory sandbox per l'IA otterranno un vantaggio competitivo – soprattutto in termini di conformità – sulla concorrenza, ma è anche vero che i diversi Stati membri, nell'istituire le rispettive regulatory sandboxes per l'IA a livello nazionale, potrebbero adottare criteri di test molto eterogenei con un conseguente rischio di frammentazione del mercato unico dell'Unione Europea.

Tuttavia, al di là delle criticità, è bene notare che le regulatory sandboxes non sono soltanto – come si è visto – una misura a sostegno dell'innovazione, ma offrono notevoli vantaggi in una prospettiva di ethics by design. Infatti, nell'ambito di una regulatory sandbox per l'IA, è possibile affrontare le questioni etiche attraverso cinque passaggi delineati da Marianne Boenink: specificare, ricostruire, sondare, ampliare e convergere.<sup>29</sup> Per prima cosa, nello sviluppo di un sistema di intelligenza artificiale, è importante *specificare* in maniera chiara quali siano le funzionalità di tale sistema, andando a definire il contesto di utilizzo, come – ad esempio – una diagnosi medica o una decisione finanziaria, gli obiettivi tecnici e quelli etici. Specificato l'oggetto (il “cosa”), ci si sofferma sul “come”. Questa seconda fase mira a *ricostruire* il modo in cui vengono affrontate questioni nodali come quelle relative ai bias nei dati con un focus sui processi decisionali e le assunzioni implicite. Segue la fase in cui si vanno a *sondare* le motivazioni, ovvero il “perché” determinate scelte sono state intraprese in un modo piuttosto che in un altro. Il quarto passaggio mira ad *ampliare* la prospettiva, coinvolgendo stakeholders esterni – per esempio utenti finali o esperti di etica – per raccogliere feedback riguardanti possibili questioni trascurate o per garantire una maggiore inclusività. Infine, si tratta di far *convergere* tutto ciò che è emerso nelle fasi precedenti, andando ad armonizzare l'istanza tecnica e quella etica con eventuali modifiche grazie ai feedback ricevuti.

Nella prima fase, come si è detto, è bene delineare gli obiettivi tecnici (ad esempio la robustezza) e quelli etici. Per quanto riguarda questi ultimi, i principi delineati da Tom Beauchamp e James Childress – beneficenza, non maleficenza, autonomia e giustizia<sup>30</sup> – rappresentano un punto di partenza utile che, recentemente, Luciano Floridi ha ampliato e adattato al contesto specifico dell'intelligenza artificiale, introducendo un quinto principio fondamentale: l'esplicabilità.<sup>31</sup> Di fronte a una proliferazione di linee guida per uno sviluppo etico dell'intelligenza artificiale – come dimostra il progetto *AI Ethics Guidelines Global Inventory* di Algorithm Watch,<sup>32</sup> che, ad aprile 2020, ne ha censite ben 167 –, Floridi propone, appunto, un quadro unificato:

Ciascuno di questi principi (i.e. beneficenza, non maleficenza, autonomia e giustizia) coglie qualcosa di apparentemente nuovo riguardo all'IA come forma dell'agire: il suo funzionamento è spesso invisibile o incomprensibile a tutti tranne (nella migliore delle ipotesi) agli osservatori più esperti. L'aggiunta del principio di “esplicabilità”, che include sia il senso epistemologico di “intelligibilità” sia il senso etico di

<sup>28</sup> Cfr. T. Madiaga, A.L. Van De Pol, “Artificial intelligence act and regulatory sandboxes”, *European Parliamentary Research Service* 2022, p. 3.

<[https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS\\_BRI\(2022\)733544\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf)>.

<sup>29</sup> Cfr. M. Boenink, *The Multiple Practices of Doing “Ethics in the Laboratory”*: A Mid-level Perspective, in S. van der Burg, T. Swierstra (eds.), *Ethics on the Laboratory Floor*, Palgrave Macmillan, London, 2013.

<sup>30</sup> T.L. Beauchamp, J.F. Childress, *Principles of Biomedical Ethics*, Oxford University Press, Oxford, 2019<sup>8</sup>.

<sup>31</sup> L. Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Cortina, Milano, 2022. Si veda anche L. Floridi, J. Cowls, “A Unified Framework of Five Principles for AI in Society”, *Harvard Data Science Review* 1, 1 (2019) <<https://doi.org/10.1162/99608f92.8cd550d1>>.

<sup>32</sup> Algorithm Watch, *AI Ethics Guidelines Global Inventory*, aprile 2020 <<https://inventory.algorithmwatch.org/>>.



“responsabilità”, è il cruciale pezzo mancante del puzzle etico dell’IA. Questo principio completa gli altri quattro: affinché l’IA sia benefica e non malefica, dobbiamo essere in grado di comprendere il bene o il danno che sta effettivamente facendo alla società e in quali modi; affinché l’IA promuova e non limiti l’autonomia umana, la nostra “decisione su chi dovrebbe decidere” deve essere informata dalla conoscenza di come l’IA agirebbe al nostro posto e, in tal caso, di come migliorare le sue prestazioni; e, affinché l’IA sia giusta, dobbiamo sapere chi ritenere eticamente o legalmente responsabile in caso di un esito grave e negativo, il che richiederebbe a sua volta un’adeguata comprensione del perché tale esito si sia prodotto.<sup>33</sup>

In tale contesto, le regulatory sandboxes offrono una “zona franca” dove testare sistemi di intelligenza artificiale e integrare il principio della esplicabilità con i quattro principi etici tradizionali. È fondamentale, infatti, bilanciare l’innovazione con il rispetto dei diritti umani fondamentali in modo tale che nessuno possa «sottrarsi alla propria responsabilità morale e giuridica affermando di non avere idea di come funziona il loro sistema e di come reagirà».<sup>34</sup> Si tratta di parole scritte da filosofi, giuristi e ingegneri in una lettera aperta pubblicata dopo il drammatico episodio del suicidio di un giovane belga, avvenuto in seguito a prolungate conversazioni con un chatbot che lo avrebbe convinto a togliersi la vita.<sup>35</sup> Questo è soltanto uno dei numerosi casi di cronaca che mostrano quanto i sistemi di intelligenza artificiale possano compromettere l’autonomia umana. La risposta dell’Europa a queste sfide è rappresentata dall’AI Act e, in particolare, dalle regulatory sandboxes, le quali consentono di affrontare le questioni etiche fin dalle fasi iniziali dello sviluppo di una tecnologia in una prospettiva di ethics by design che, grazie all’“effetto Bruxelles”,<sup>36</sup> ha il potenziale di diventare uno standard normativo globale.

<sup>33</sup> L. Floridi, *Etica dell’intelligenza artificiale. Sviluppi, opportunità, sfide*, cit., pp. 100-101.

<sup>34</sup> N.A. Smuha, M. De Ketelaere, M. Coeckelbergh, P. Dewitte, Y. Pouillet, *Open Letter: We are not ready for manipulative AI – urgent need for action*, KU Leuven, 31 marzo 2023 <<https://www.law.kuleuven.be/ai-summer-school/open-brief/open-letter-manipulative-ai>> (traduzione mia).

<sup>35</sup> Si veda P.-F. Lovens, “Sans ces conversations avec le chatbot Eliza, mon mari serait toujours là”, *La Libre Belgique* 28 (marzo 2023) <<https://www.lalibre.be/belgique/societe/2023/03/28/sans-ces-conversations-avec-le-chatbot-eliza-mon-mari-serait-toujours-la-LVSLWPC5WRDX7J2RCHNWPST24/>>; “Le chatbot Eliza a brisé une vie : il est temps d’agir face à l’IA manipulatrice”, *La Libre Belgique* 29 marzo 2023 <<https://www.lalibre.be/debats/2023/03/29/le-chatbot-eliza-a-brise-une-vie-il-est-temps-dagir-face-a-lia-manipulatrice-BSGGRV7IBRDNROO33EWGFVMWAA/>>.

<sup>36</sup> Su questo cfr. A. Bradford, *Effetto Bruxelles. Come l’Unione Europea regola il mondo*, Franco Angeli, Milano, 2021.