

OLTRE LA *PRIVACY* INFORMATIZIONALE: LIBERTÀ DI SCELTA E DI IDENTITÀ NELL'ERA DELLA PROFILAZIONE ALGORITMICA

Simona Tiribelli*

Abstract

The protection of individual identity is one of the most prominent and discussed issues within the debate in the field of ethics of artificial intelligence (AI) and ethics of algorithms specifically. However, thus far, the debate has mainly focused on the threats raised by AI and algorithm-based ICTs to personal identity in terms of informational privacy infringements, while no specific contribution has been developed on their impact on individual moral identity and, in particular, on our freedom to develop an authentic moral identity. This paper aims at filling this gap. To do so, by drawing on moral philosophy, we pursue an ethical inquiry and show how algorithm-based ICTs can negatively affect the preconditions of our freedom of choice and action and, by doing so, erode over time our freedom to develop authentic moral identities. Finally, we pinpoint key ethical criteria that acknowledge, but also go beyond existing informational privacy-preserving solutions, in order to design algorithm-based ICTs that can ensure at least at a minimum threshold the protection of our freedom to choose, act, and therefore, to develop authentic moral identity in our hyper-profiled *onlife* environments.

Keywords: Artificial Intelligence, Algorithms, Autonomy, Choice, Identity

Introduzione

L'applicazione pervasiva di tecnologie digitali di informazione e comunicazione (ICT) sempre più basate su sistemi di intelligenza artificiale (IA) quali, in particolare, gli algoritmi di apprendimento automatico e profondo, ormai in quasi ogni ambito della vita umana – dalle innumerevoli applicazioni nel vasto mondo dell'*Internet of things* (IoT) ai fornitori di servizi *online* (*online service providers*: OSP), quali i motori di ricerca Internet,

* Università di Macerata.

gli aggregatori di notizie e i *social networking services* (SNS) – ha reso la questione della tutela dell'identità individuale, principalmente recepita dal dibattito come protezione della *privacy informazionale* dell'individuo (Blume 2002; Steeves, Lucock e Keer 2009; Nissenbaum 2010; Ess 2010; Polonetsky e Tene 2013; Hull 2015; Floridi 2016), uno dei temi centrali negli ambiti dell'etica e dell'intelligenza artificiale (IA) e, in particolare, dell'etica degli algoritmi (Mittelstadt et al. 2016; Tsamados et al. 2022). A segnare l'urgenza del tema è stato soprattutto lo scandalo di Cambridge Analytica (Rosenberg 2018), che ha visto lo sfruttamento intenzionale dei sistemi algoritmici che regolano i nostri più comuni OSP e SNS per la profilazione e l'acquisizione di informazioni personali di milioni di utenti con lo scopo di influenzarne le scelte di voto e, nello specifico, le relative identità socio-politiche, per soddisfare obiettivi e interessi di parti terze. Le prime preoccupazioni etiche sulle capacità delle ICT basate su algoritmi di inferire caratteristiche, abitudini, preferenze e interessi degli utenti erano infatti confinate alla sfera commerciale e alle scelte o al comportamento di acquisto *online* degli utenti; dunque, erano esclusivamente rivolte alle tecniche di profilazione e *targeting* di tipo pubblicitario a fini commerciali (Richards 2013). Di conseguenza, la ricerca e le soluzioni riguardanti l'utilizzo delle ICT algoritmiche si sono principalmente incentrate sulla tutela delle scelte e delle identità degli utenti come consumatori attraverso aggiornamenti delle regolamentazioni e normative sulla *privacy* (Calo 2014) e delle tecniche di protezione dei dati (Turow 2011). Il caso di Cambridge Analytica ha invece esteso il portato e le preoccupazioni etiche relative all'impiego delle tecnologie algoritmiche sulle enormi quantità di dati oggi disponibili sugli individui ben oltre la sola sfera commerciale. Lo sfruttamento intenzionalmente dannoso dei sistemi algoritmici al fine di influenzare sia le scelte sia le identità sociali e politiche degli utenti ha posto in luce su come sia qualcosa di più profondo a essere in pericolo: la libertà decisionale, sia di scelta sia di azione degli individui e – aggiungiamo –, con esse, la libertà di sviluppare identità non solo socio-politiche, ma anche morali *autentiche*, ovvero informate da ragioni, valori, credenze e progetti morali fondativi, sviluppati e avallati in modo libero nel tempo.

A livello istituzionale, un esempio di questo riconoscimento lo si trova nell'adozione da parte del Comitato dei Ministri dell'Unione Europea (UE), nel febbraio del 2019, della *Dichiarazione* sulle pericolose e invisibili capacità manipolative degli algoritmi, in cui l'UE invita gli Stati Membri a una maggiore vigilanza sull'uso dei dati personali degli utenti – dagli stessi di sovente divulgati per premi relativamente piccoli di convenienza personale – poiché utilizzati per alimentare e allenare tecnologie dotate del-

la capacità di *prevedere e rimodellare* le preferenze degli utenti. Nel mondo accademico, il giurista Brett Frischmann e il filosofo Evan Selinger descrivono questo potere di influenza detenuto dalle tecnologie algoritmiche tramite il concetto di “ingegneria tecno-sociale”, riferendosi con questo ai quei processi tramite cui tali tecnologie, monitorando a livello micro, meso e macro le nostre vite e utilizzando i dati derivanti da tale monitoraggio continuo per costruire ambienti “tecno-sociali” in cui “technologies and social forces align and impact on how we think, perceive, and act” (Frischmann e Selinger 2018, p. 4), riplasmano profondamente come pensiamo e conduciamo le nostre vite, poiché inseriti in spazi progettati per renderci prevedibili e riprogrammabili. In modo simile, l'accademica e scrittrice Shoshana Zuboff, ne *Il Capitalismo della Sorveglianza* (2019), denuncia gli algoritmi come forze che trattano l'esperienza umana quotidiana in modo strumentale e come materia prima gratuita da cui inferire e ricavare prodotti dotati di alto potere predittivo tramite cui poi indurre specifiche modifiche comportamentali. Dal caso in questione, il corpo di letteratura sul tema è aumentato esponenzialmente e si distingue al suo interno tra coloro che argomentano a favore della definizione dei sistemi algoritmici come strumenti di manipolazione pubblica (Susser, Roessler e Nissenbaum 2019) e coloro che ne intravedono un potenziale esclusivamente persuasivo (Williams 2018). Tuttavia, poche sono le riflessioni, soprattutto in ambito filosofico-morale, che esaminano criticamente l'impatto e le conseguenze etiche dei sistemi algoritmici sulla formazione dell'identità personale e, nello specifico, sulla formazione della nostra identità morale, quale frutto dinamico di scelte, azioni e comportamenti sviluppati, criticamente vagliati e consolidati liberamente nel tempo. Queste riflessioni riconducono principalmente le sfide all'identità sollevate dagli algoritmi a violazioni della *privacy* degli individui e l'identità è innanzitutto intesa e concettualizzata in termini informazionali (Floridi 2006): l'idea di base è che noi siamo costituiti dalle nostre informazioni e che, dunque, violare le nostre informazioni (violazione della *privacy*) significa violare la nostra identità (si pensi, ad esempio, ai fenomeni di *data breach* quali divulgazioni accidentali o illecite di dati).

Sebbene tali riflessioni apportino un contributo fondamentale alla comprensione, nelle nostre società odierne, dell'identità personale come strettamente connessa a questioni di *privacy* individuale (Floridi 2005, 2016), le soluzioni per la sua tutela prospettate finora, da sole, non si sono dimostrate sufficienti alla protezione dell'identità personale nelle società algoritmiche contemporanee. Infatti, la possibilità che vi sia uno spazio di *privacy* o, come definito in letteratura, non soggetto a profilazione algoritmica, che

esprimerebbe il cosiddetto diritto e imperativo etico to *let be alone* (Floridi 2016; Ess 2010, p. 108), rivendicato nelle riflessioni menzionate, ovvero di uno spazio informazionale privo di interferenze algoritmiche, che ci permetta di essere “masters of our own journeys and keep our choices and identity open” (Floridi 2005, p. 310), sembra essere difficile sia da individuare sia da tutelare. Come affermato da eminenti filosofi e giuristi nell’ambito dell’etica dell’IA e degli algoritmi (Watcher e Mittelstadt 2019), l’utilizzo illecito delle tecnologie algoritmiche per catturare informazioni personali sempre più intime, anche in quei paesi dell’UE dove sono già in atto normative giuridiche aggiornate (quali il *General Data Protection Regulation*: GDPR), è molto complesso da scovare e regolamentare, a causa della fitta rete di algoritmi in cui siamo immersi e, soprattutto, della continua e sempre più variegata quantità di dati che produciamo e con cui, anche indirettamente, li alleniamo. A questo proposito, si pensi a tecniche di profilazione algoritmica come il *collaborative filtering*, che consente di inferire e scoprire informazioni predittive relative a interessi e preferenze di un insieme dato di utenti a partire da quantità indifferenziate di dati, oppure alla possibilità di profilazione per associazione tramite dati di valore apparentemente neutro che riescono a fungere da *proxy* per risalire a connotati e caratteristiche di categorie o gruppi giuridicamente definiti come protetti (ad esempio: minoranze etniche). Sebbene fondamentale, l’attuale concettualizzazione della questione della tutela dell’identità individuale quasi esclusivamente nei termini di protezione della *privacy* informazionale non sembra, da sola, e a oggi, fornirci alcuna bussola propriamente etica, e dunque ulteriore a quella giuridica esistente, su come orientare *by design* il funzionamento dei sistemi algoritmici affinché non rappresentino un limite o pericolo allo sviluppo dell’identità socio-politiche e, soprattutto, morali degli individui. In aggiunta, sebbene tali contributi riconoscano la stretta interconnessione tra libertà di scelta e di identità (Floridi 2005, Floridi e Taddeo 2018), invitando alla tutela dell’identità individuale come il diritto all’apertura delle nostre scelte, azioni e delle nostre identità, un’indagine propriamente etica, che tragga spunti dalla filosofia morale, sull’impatto delle tecnologie algoritmiche sulla nostra libertà di scelta, di azione e di identità in un ambiente ormai totalmente digitalizzato, interconnesso e/o, per utilizzare un noto neologismo attuale, “*onlife*” (Floridi 2014), è al momento ancora assente nel dibattito. Il presente lavoro tenta di colmare questa lacuna fornendo una prima concettualizzazione dell’impatto delle ICT basate su algoritmi sulla nostra libertà di scelta e di azione *onlife*, mettendone in luce, nello specifico, le implicazioni etiche che ne derivano sulla nostra libertà di sviluppare identità morali autentiche, ovvero orienta-

te e informate da valori e progetti morali fondativi criticamente sviluppati, espressi e avallati in modo libero nel tempo. La tesi che sosteniamo è che un'indagine etica sull'impatto degli algoritmi sulle condizioni che possono garantire, almeno a una soglia minima, la nostra libertà di scegliere, di agire e, nel tempo, di sviluppare identità morali autentiche possa offrire una prospettiva fruttuosa per integrare gli approcci di *privacy* informazionale attualmente sviluppati in letteratura per la tutela dell'identità individuale. In particolare, cerchiamo di porre in luce come tale indagine possa illuminare condizioni o dimensioni specifiche che richiedono di essere tutelate e che dunque possono indicare criteri di *design etico* (*ethics by design*) delle tecnologie algoritmiche al fine di preservare, almeno a una soglia minima, e/o promuovere la nostra possibilità di scegliere e agire liberamente e, così facendo, nel tempo, sviluppare autentiche identità morali.

A tal fine, il lavoro si struttura in tre sezioni.

Nella prima sezione, poniamo in luce quali siano le condizioni necessarie ma non sufficienti riconosciute – con un certo accordo – dal dibattito sviluppato in filosofia morale al fine di garantire la nostra libertà di scelta e azione a una soglia minima, illuminando, al contempo, come tali condizioni necessitino di essere ripensate nei contesti di scelta contemporanei, quelli delle nostre società *onlife*, sempre più permeate da algoritmi e, dunque, digitalizzate e/o datificate; in sintesi: nelle nostre società dell'informazione. Una volta messe in luce brevemente tali condizioni, la seconda sezione si incentra sull'analisi di come le tecnologie algoritmiche stiano impattando sulle condizioni *sine qua non* al fine di tutelare a una soglia minima la nostra libertà di scelta, azione e di formazione di una identità morale autentica. Infine, nella terza sezione, sulla base delle condizioni evidenziate e dei risultati della nostra indagine etica, indichiamo alcuni criteri complementari o integrativi, non sostitutivi, di quelli offerti dagli approcci esistenti di *privacy* informazionale a tutela dell'identità individuale, illuminando come possano rivelarsi preziosi per la progettazione di ICT algoritmiche in grado di salvaguardare, almeno a una soglia minima, la nostra libertà di scelta, di azione e di formazione di identità morali autentiche nell'era della profilazione algoritmica.

1. *Libertà di scelta, azione e identità onlife*

Al fine di indagare se e come le tecnologie algoritmiche stiano inficiando la nostra libertà di scelta, di azione e di identità, è preliminare tentare di mettere in luce *quali* siano le condizioni che rendono possibile la nostra li-

bertà di scelta e azione, perlomeno a una soglia minima, e soprattutto come oggi necessitano di essere ripensate alla luce dell'enorme processo di da-
tificazione che interessa, senza esclusione, ogni ambito della nostra realtà.

Se siamo liberi nella scelta è una questione emblematica che interroga la tradizione filosofica fin dalle sue origini. Sebbene siano stati molteplici i tentativi filosofici e scientifici di liquidarla, dalle neuroscienze alla fisica fino alla biologia, inaugurando controversie secolari relative alla sua natura o *statuto metafisico* (si consideri il dibattito filosofico sulla libertà di scelta come libertà della volontà o libero arbitrio), l'esercizio della libertà di scelta e di azione, ovvero la libertà di scelta nella sua *valenza pratica* quale possibilità dell'individuo di scegliere e agire liberamente a) dato un contesto o disponibilità di *opzioni alternative* e b) secondo motivi di cui può dirsi *in controllo*, si è mostrata resiliente ai plurali condizionamenti e ostacoli che molti filosofi (e non solo) le hanno contrapposto nel tempo, quali, solo per menzionarne i principali, la prescienza divina, il determinismo fisico e gli innumerevoli condizionamenti socio-ambientali, neurofisiologici e genetici.¹

Sebbene nella considerazione della libertà e, nello specifico, nei concetti di scelta e azione, si intreccino ambo le dimensioni, metafisica (*natura della libertà*) e pratica (*esercizio della libertà*), è esclusivamente a quest'ultima a cui il presente lavoro intende riferirsi: alla possibilità di un esercizio libero della scelta, il cui darsi non è solo imprescindibile per l'attribuzione di responsabilità,² oltre che di colpa e biasimo, ma anche – e di particolare interesse in questa sede – per lo sviluppo di una identità morale autentica, ovvero di un'identità che riflette ragioni, valori e progetti morali scelti e, dunque, avallati in modo libero nel tempo. Sebbene la tradizione filosofica, soprattutto nell'ambito morale, abbia spesso trattato le due dimensioni in modo congiunto, facendo dipendere l'esercizio di scelte e azioni libere dallo statuto libero o non libero della volontà, è possibile trovare delle eccezioni, oltre che in alcuni scritti di filosofi eminenti quali *On Liberty and Necessity* (1654) di Thomas Hobbes, o *Freedom of Will and Freedom of Action* (1985) di Rogers Albritton, già nel pensiero di Immanuel Kant, che distingueva in modo preciso fin dalla *Critica della Ragion Pura* (1787) la

1 Per un approfondimento del dibattito su libertà di scelta e determinismo (determinismi) e, nello specifico, per un'ampia, dettagliata e sistematica trattazione delle teorie che difendono l'idea che gli esseri umani siano liberi nonostante le teorie contrarie, si veda De Caro (2004). Per una ricostruzione storica del dibattito, si veda Mori 2001.

2 Sull'idea di libertà quale condizione di possibilità (*ratio essendi*) della responsabilità, necessario è il rimando alla *Critica della Ragione Pratica* di Kant (1788).

concezione di libertà definita come trascendentale, o libero arbitrio *stricto sensu*, e una concezione di libertà pratica, conoscibile tramite i sensi e l'esperienza sensibile, come libertà di scegliere e agire – o dirigere la propria condotta – per amore di beni più lontani e sensibili. Senza pretesa di ricostruire il dibattito sulla libertà di scelta in ambito filosofico morale, sia per limiti di spazio, sia poiché ci condurrebbe lontano dal nostro scopo, ovvero l'analisi dell'impatto dei sistemi di IA e degli algoritmi sulla libertà di scelta, azione e identità, precisiamo che non tratteremo, dunque, della natura intrinseca della libertà di scelta (libero arbitrio) che tende a riferirsi al concetto spesso vago o enigmatico di volontà, bensì del suo esercizio e, dunque, delle condizioni necessarie ma non sufficienti (*sine qua non*) al fine di poterne garantire la possibilità, soprattutto nelle nostre società contemporanee, in cui gli ambienti di scelta sono sempre più permeati se non governati da algoritmi.

Nel dibattito sviluppato nell'ambito della filosofia morale sulla libertà di scelta è possibile rintracciare da Cartesio in poi, non senza disaccordi, una certa convergenza su almeno due condizioni congiuntamente necessarie o *sine qua non* (De Caro 2004; Mori 2001) all'esercizio libero della scelta: la *disponibilità di opzioni alternative* e l'*autonomia* o principio di autodeterminazione. In altre parole, una scelta può dirsi davvero libera *se e solo se* a) l'agente può agire diversamente da ciò che *de facto* fa e *se e solo se* b) i motivi e le ragioni delle sue scelte risultano in suo controllo, ovvero non siano il frutto di scelte casuali o di coercizione (De Caro 2004, p. 4). È chiaro che se un agente è libero di eseguire una certa scelta e azione, sarà anche libero di non eseguirla. Allo stesso tempo, è anche intuitivamente chiaro che per garantire questa libertà, la scelta tra potenziali corsi di azione non può essere casuale, vale a dire, non può derivare da fattori completamente fuori controllo dell'agente (come quando, per esempio, un agente decide se compiere una certa azione sulla base del lancio di una moneta, o quando è costretto a compiere una certa scelta), quanto piuttosto il risultato di ragioni e motivi di cui può dirsi in controllo. Al di là delle sfumature concettuali che le condizioni messe in luce acquisiscono nell'ampio dibattito filosofico sulla libertà di scelta, queste rappresentano il minimo comune denominatore di molte tradizioni filosofiche, anche all'interno di un altro dibattito sulla libertà di scelta, che non analizzeremo in questa sede, di natura socio-politica (Skinner 2008, 2012). Per lo scopo della nostra analisi, tali condizioni ci forniscono un primo terreno di valutazione etica per comprendere l'impatto delle ICT algoritmiche sulla nostra libertà di scelta, azione e, nel tempo, di sviluppo di identità morali autentiche. Nella prossima sezione, infatti, vagliamo se le ICT stiano minando la nostra libertà

di scelta, azione e identità creando nuovi vincoli o condizionamenti sulle condizioni evidenziate; in altre parole, indaghiamo se e come tali tecnologie interagiscono, abilitando o disabilitando, promuovendo o ostacolando, le condizioni *sine qua non* sopra menzionate. Prima di inoltrarci nell'analisi etica in questione, tuttavia, è necessario interrogarsi sulla necessità di ripensare o perlomeno ricontestualizzare le condizioni sopra citate negli ambienti in cui compiamo le nostre scelte, le nostre azioni e, con esse, sviluppiamo le nostre identità sociali, politiche e morali; ambienti sempre più digitalizzati, datificati, dunque iper-connessi, ibridi o *onlife*: ambienti sempre più concepibili in termini informativi.

In effetti, con l'avvento della quarta rivoluzione (Floridi 2014), o rivoluzione digitale, ovvero della crescente ibridazione o digitalizzazione del mondo contemporaneo, è ormai quasi inconfutabile l'idea secondo cui viviamo completamente immersi in un ambiente informativo: un ambiente in cui incessantemente ci nutriamo in *input* ed elaboriamo e riproduciamo informazione in *output*. Tutto – dai nostri nomi e caratteristiche psico-fisiche fino alle nostre più intime preferenze, valori, abitudini, opinioni, credenze, affiliazioni, interazioni, orientamenti – non è più solo teoricamente (o idealmente) concepibile e descrivibile in termini informativi, ma è concretamente codificabile in dati (processo di *datafication*) in cui osservare e, soprattutto, da cui estrarre enormi quantità d'informazioni sensibili sulle persone, sulle loro più profonde connessioni, e sul mondo, nella sua generalità e complessità, a livello micro e macro. Si pensi a quanto le nostre cronologie di ricerca datifichino i nostri pensieri o interessi in semplici *queries*, quanto i luoghi che visitiamo sia *online* (es. siti web) che *offline* (luoghi fisici) digitalizzino le nostre abitudini, stili di vita, frequentazioni, ma anche forniscano dati sul nostro stato sociale, economico e persino medico. A proposito di quest'ultimo, dovrebbe sorprenderci che solo di recente, lo scorso luglio 2022, a seguito del dibattito post sentenza della corte suprema *Roe vs. Wade* sull'aborto, Google ha annunciato che cancellerà i dati di geolocalizzazione delle utenti (specifica, quella del genere femminile, non di poco rilievo a fini di profilazione) che si recheranno in ospedali, cliniche abortive o di fertilità. Questo annuncio è emblematico di quanto non solo ciò che è accaduto e che è accaduto, ma anche ciò che sta per o potrebbe accadere, ovvero il *range* delle nostre possibilità, dei nostri possibili corsi di azione futuri, sia datificabile. In sintesi, la diffusione e l'uso pervasivo delle ICT algoritmiche, insieme alla nostra presenza totalizzante nei SNS, permettono di pensare e ri-concettualizzare l'ambiente iperconnesso o *onlife* che abitiamo in termini informativi. Senza inoltrarci nella possibilità di una modifica ontologica dell'ordine delle cose o dell'essenza

del mondo, come avanzato da alcuni rinomati filosofi nel settore (Floridi 2005), analisi che ci ricondurrebbe verso un dibattito di stampo più metafisico che pratico sul reale e sulla libertà di scelta, è sufficiente questa riconcettualizzazione informazionale del modo in cui pensiamo agli ambienti in cui siamo immersi per comprendere la necessità di ripensare anche ai contesti e alle condizioni che rendono possibile la nostra libertà di scelta secondo una prospettiva informazionale. Quivi, riconosciamo e prendiamo le mosse dagli apporti delle teorie contemporanee sull'identità informazionale (Floridi 2016), poiché cruciali, come anticipato nella sezione introduttiva, alla comprensione del ruolo centrale dell'informazione negli ambienti iperconnessi contemporanei, per illuminarne le implicazioni ulteriori sulle condizioni necessarie all'esercizio libero delle nostre scelte.

Secondo questa prospettiva, l'esercizio della nostra autonomia non risiede solo nel controllo delle nostre informazioni (Floridi 2016), controllo – come affermato sopra – oggi sempre più difficile da esercitare, sia da parte delle istituzioni sia dagli utenti; quanto nella possibilità di avallare criticamente ciò che a livello informazionale – quelle informazioni (o opzioni informazionali) – può dare origine o costituirsi come potenziale motivo delle nostre scelte. Questo poiché ogni aspetto e componente di noi stessi e del mondo, infatti, può essere datificato come informazione, e dunque ogni informazione trasmette e incorpora, oltre che un dato descrittivo, uno specifico portato etico normativo di valori, orizzonti di significato, credenze e preferenze, che apre a diverse possibilità e scelte e, dunque, a corsi di azione alternativi e, nel tempo, a percorsi di formazione identitaria (socio-culturale, politica e, in particolare, morale) distinti. Ciò significa anche riconcettualizzare, oltre che l'autonomia (o principio di autodeterminazione), l'altra condizione identificata per garantire a una soglia minima la nostra libertà di scelta, di azione e, nel tempo, di identità: la disponibilità di opzioni alternative. La nostra disponibilità di opzioni alternative, infatti, può essere ripensata, alla luce di un ambiente ibrido che ingloba la dimensione dell'*online* o cyber-spazio, come disponibilità *onlife* di opzioni informazionali. In altre parole: se tutto può essere considerato, incorporato o inglobato come informazione o tradotto in termini informazionali, ogni dato informazionale può costituire una opzione che informa o sui cui l'individuo può agire nel processo decisionale: ogni dato informazionale può costituire una opzione alternativa, ovvero potenzialmente dotata del portato sia descrittivo sia normativo per aprire a corsi di azioni alternativi. Possono essere dunque considerate opzioni tutte quelle informazioni (su noi stessi, il mondo e gli altri) che informano e sono disponibili al soggetto nel processo decisionale. In sintesi: possono essere concepibili

come opzioni quelle informazioni che sia costituiscono sia incorporano – nel processo decisionale – possibili corsi di azione alternativi, ponendo l'individuo nello stato di a) poter *fare altrimenti*.

Gli ambienti che abitiamo offrono potenzialmente una serie infinita di opzioni informazionali. Tuttavia, a seconda dei contesti sociali e ambientali in cui siamo inseriti, possiamo accedere ad alcune piuttosto che ad altre: in sintesi, a contare è l'effettiva disponibilità delle opzioni informazionali a cui possiamo accedere o meno e, a fini di una garanzia minima di libertà di scelta, di azione e di identità, un livello sufficiente di eterogeneità informazionale tale da garantire l'apertura delle scelte e, dunque, a corsi di azione realmente alternativi.³ È interessante accennare a come le prime riflessioni sui media digitali avevano posto particolare enfasi sulla possibilità di godere, negli spazi *online* (o cyberspazio), di una disponibilità di scelte e opzioni e di un grado di libertà di espressione e formazione identitaria diametralmente maggiore rispetto a quello fruibile negli spazi analogici (Bakardjieva e Gaden 2012). Tali riflessioni sottolineavano come risiedesse nelle libertà o opzioni di scelta illimitate offerte⁴ dagli ambienti *online* (il noto motto: nel *cyberspazio* puoi essere e diventare chiunque tu voglia) – resa in realtà possibile dalla loro iniziale a-regolamentazione – risiedesse nel loro rapido sviluppo e nella forte presa sugli individui. Pertanto, l'idea diffusa era che la compenetrazione e ibridazione tra spazi analogici, fisici o incarnati (*offline*), e digitali o virtuali (*online*), negli ambienti di cui disponiamoci oggi, per l'appunto *onlife*, avrebbe aperto a scenari e soprattutto a contesti di scelta in cui vedere notevolmente potenziata la nostra disponibilità o *range* di opzioni alternative – a livello sia quantitativo, sia e soprattutto qualitativo (eterogeneità) – e, con esse, la nostra autonomia, autorialità o possibilità di autodeterminazione attraverso lo sviluppo di percorsi identitari molteplici e inediti; in sintesi: spazi nuovi, di promozione della libertà di scelta, di azione e di identità. Tuttavia, le cose sembrano presentarsi diversamente, sia alla comune percezione empirica collettiva, sia sulla base dei fenomeni o pericoli illuminati dagli studi evidenziati in *ouverture* a questo contributo.

3 In altri termini: la nostra libertà di scelta tra alternative non è mai assoluta, ma dipende sempre da certe informazioni disponibili (nonché da altri fattori, come, ad esempio, l'educazione, ecc.); in questo senso, i nostri *range* di opzioni di scelta sono sempre – almeno parzialmente – predefiniti: scegliamo in contesti in cui siamo sempre orientati a considerare una serie di opzioni alternative come effettivamente disponibili, piuttosto che altre (ringrazio un *referee* anonimo per la richiesta di esplicitazione).

4 Aggiungiamo: offerta (informazionale) quantomeno più ampia, se comparata a quella ad esempio delle ICT tradizionali, considerata l'assenza di limiti temporali e, in una certa misura, spaziali, che connota gli ambienti *online* o il cyberspazio.

Come mostriamo nella prossima sezione, negli ambienti (o contesti di scelta) informazionali, in cui prepariamo e compiamo le nostre scelte e dove sviluppiamo, nel tempo, le nostre identità, quali ambienti caratterizzati da disponibilità *onlife* di opzioni informazionali, la fruibilità (o disponibilità) e l'eterogeneità delle opzioni accessibili al soggetto non sono oggi solo influenzate da fattori o agenti sociali, culturali, geografici e/o politici, ma anche da forme di agenzia completamente nuove o da agenti artificiali, che non solo possono creare una illusione di maggiore libertà di scelta, azione e di identità, grazie all'uso combinato di tecniche di profilazione e personalizzazione dei contesti di scelta *onlife*, ma rischiano di predeterminarli in modi che rischiano di erodere profondamente la nostra autonomia sulle nostre scelte e azioni e, con esse, di minare la nostra autorialità – e autenticità – nello sviluppo identitario. È all'indagine di questa azione algoritmica sulle condizioni che abbiamo posto in luce come *sine qua non* della nostra libertà di scelta, azione e identità nei contesti informazionali delle nostre società contemporanee che dedichiamo la sezione a seguire. Nello specifico, non indagheremo la natura degli agenti algoritmici, ovvero, se siano o meno agenti morali, a cui è stato dedicato un corpo consistente di letteratura nell'ambito dell'etica dell'IA e degli algoritmi (Liao 2020); approfondiremo, invece, valutandone al contempo la relativa bontà, il ruolo di influenza o *governance* che le tecnologie algoritmiche esercitano prima a) sulle opzioni a noi disponibili a livello informazionale nei contesti di scelta *onlife*; poi b) sull'autonomia che possiamo esercitare su ciò che – a livello informazionale: credenze, valori, ragioni, progetti morali – informa e motiva le nostre scelte, azioni e comportamenti e, nel tempo, definisce lo sviluppo delle nostre identità sociali, politiche, culturali e, in particolare, delle nostre identità morali.

2. *Intelligenza artificiale e libertà di scelta, azione e identità: nuovi vincoli epistemologici e pratici*

Per decomprimere il crescente ruolo di *governance* svolto dagli algoritmi nei contesti di scelta *onlife* contemporanei e, soprattutto, illuminarne il *modus operandi*, soprattutto in relazione alle condizioni *sine qua non* della libertà di scelta sopra evidenziate, è sufficiente considerarne il loro ruolo proprio rispetto a quel processo di *datafication* che caratterizza oggi strutturalmente le società contemporanee e che, nel digitalizzare (quasi) ogni aspetto o elemento del reale, ne permette una sua ricompressione informazionale.

Questo processo di *datafication*, secondo cui tutto può essere tradotto in dati e, dunque, essere oggetto di analisi, processazione e semanticizzazione (produzione di informazione e conoscenza), è infatti governato da algoritmi: modelli costituiti da segmenti di codice matematico e incessantemente alimentati da quei dati che direttamente cediamo (*provided data*), spuntando la clausola del consenso informato, ad esempio, quando accediamo o fruiamo di determinati OSP o SNS; da tracce *onlife* che lasciamo indirettamente (*observable data*) interagendo con elementi digitali, dati relativi alle nostre ricerche *online* come i siti web che consultiamo, alle nostre interazioni sia con gli altri che con l'IoT, ai nostri spostamenti e acquisti *online* e *offline*; e infine, da quei dati *derivabili* e/o *inferibili*, ovvero veri e propri nuovi set di dati e di informazioni tratte dalla correlazione (inclusa analisi probabilistica) dei dati già raccolti, precedentemente menzionati, con altri enormi set di dati ampiamente disponibili, aggregati e disaggregati. Le ultime due categorie di dati (osservati, inferiti o dedotti) comportano un basso livello di consapevolezza degli individui interessati, appunto poiché trattati o creati senza alcun coinvolgimento diretto degli stessi. A capo di questo trattamento, processazione e semanticizzazione troviamo appunto diverse tecniche algoritmiche. Infatti, sono oggi molteplici e diversi sia i modelli sia le tecniche algoritmiche che vengono impiegate per governare la fitta rete di ICT digitali dispiegate nelle nostre società informazionali e sulle massive quantità di dati disponibili. Tra questi, gli algoritmi di *machine learning* (ML) detengono un importante primato, proprio in virtù della loro natura probabilistica, che permette loro di sviluppare e dotarsi di un potenziale e/o capacità predittiva. Infatti, a differenza dei modelli statistici o deterministici – che al fine di soddisfare un obiettivo o svolgere un compito prestabilito sono progettati per seguire una serie predefinita di regole o istruzioni –, gli algoritmi di ML sono progettati per imparare in modo probabilistico *come* raggiungere un compito o obiettivo predefinito senza che vi siano percorsi causali predefiniti. Nel caso dei modelli probabilistici, infatti, gli *output* non sono il risultato di una serie di istruzioni algoritmiche, ma vengono dedotti e inferiti direttamente dai dati. Questo è reso possibile dall'impiego di diverse tecniche algoritmiche con *task* differenziati, il cui comune denominatore è la processazione di grandi set di dati, di solito disaggregati, e lo sviluppo di innumerevoli e rapidi tentativi (per errore) – nella forma di scommesse probabilistiche – volti a inferire o scoprire quali modelli (*patterns*), nodi (attribuzione di pesi) e correlazioni sono *di valore* per l'obiettivo predefinito, ovvero permettono loro di svolgere nel modo più efficiente il compito/obiettivo pre-stabilito, qualunque esso sia – sebbene, nelle ICT e nei SNS, questo tenda a coincidere con la massimizzazione del click di ciò che in-

crementa gli utili (acquisti, film, pubblicità, abbonamenti, solo per citarne alcuni) dei *providers* tecnologici che si celano dietro alle piattaforme o ai servizi in questione.

Tali tecniche algoritmiche sono definite nello specifico tecniche di scoperta della conoscenza (*knowledge-discovery techniques*) e, tra queste, la profilazione algoritmica a scopi predittivi – o profilazione predittiva – è tra le più pervasive. In breve, tale tecnica agisce scalando enormi quantità di dati per generare prodotti di previsione comportamentale che si basano su processi di etichettatura e categorizzazione dell’utenza nel suo insieme. Scalando i dati disponibili, le tecniche in questione permettono di estrarre (*data mining*) associazioni, affinità, caratteristiche su grandi numeri di utenti di valore, poiché ne accomunano il comportamento di scelta, sviluppando così prodotti predittivi nella forma di profili identitario-comportamentali tramite cui etichettare gli individui e poi raggrupparli in categorie di utenza a cui presentare contenuti/contesti informazionali specifici, ovvero personalizzati, designati al fine di raggiungere la *task* predefinito (ad esempio: come portare una certa utenza verso l’acquisto di certi prodotti nel modo più efficiente). I profili assegnati dagli algoritmi agli utenti non sono dunque altro che prodotti di previsione – per associazione o affinità – delle nostre identità culturali, socio-politiche e morali, a loro volta dedotte a partire da quelle caratteristiche, preferenze, vulnerabilità, credenze e valori (solo per citarne alcuni) inferibili dal nostro “stare al mondo *online*”: prodotti che dovrebbero predire, date certe caratteristiche, come è probabile che ci comporteremo, sceglieremo e agiremo in futuro.

In tal senso, è possibile mettere in luce un duplice ruolo di *governance* svolto dagli algoritmi: tali sistemi decidono – nella forma di scommesse probabilistiche – quale sia l’identità dell’individuo in questione e, sulla base di questo profilo, determinano i contenuti informazionali che saranno a lui disponibili, come specifico utente, parte di un gruppo determinato algoritmicamente. In altri termini: ubiqui algoritmi di profilazione decidono come identificarci e categorizzarci, governando il “modo in cui siamo visti”, dove il profilo assegnato all’utente tende ad assumere la forma di un aggregato di macro-caratteristiche che lo descrivono in relazione ad altri utenti, come caratteristiche comuni tra utenti, poiché sviluppato in modo funzionale al raggiungimento di obiettivi di parti terze; obiettivi che spesso non coincidono con quelli delle persone/utenti in questione. Per comprendere il portato di questi algoritmi basti pensare a come, sulla base di tali profili, a un particolare utente può essere escluso l’accesso a una determinata università o posizione lavorativa, a un programma di cure speciali o a tasso di mutuo agevolato, alla visualizzazione di un’offerta di lavoro *online*

oppure a un'assistenza sanitaria speciale – e tutto ciò sulla base di ragioni spesso precluse, per proprietà commerciale, o complessità algoritmica, agli interessati (Pasquale 2015). In altre parole, è difficile che l'individuo conosca o gli sia rivelato quali informazioni (correlazioni o nodi) conducano all'elaborazione del suo profilo e/o alla sua categorizzazione.

Se gli algoritmi di profilazione governano il modo in cui siamo visti, descritti e categorizzati a livello identitario e comportamentale a fine predittivo, le tecniche algoritmiche di personalizzazione informazionale, che includono, degni di menzione, gli algoritmi di classificazione, di filtro e i sistemi di raccomandazione, agiscono sui nostri ambienti o contesti informazionali, filtrandoli sulla base dei profili inferiti e assegnati, e così facendo governando il modo in cui percepiamo la realtà, gli altri e (persino) noi stessi. In questo senso, gli algoritmi non solo governano il “modo in cui siamo visti”, ma agendo sulle informazioni a noi disponibili, in altre parole, sulla nostra disponibilità di opzioni informazionali, e decidendo quante e quali informazioni sono a noi disponibili e per noi più rilevanti sulla base dei profili sviluppati, governano il nostro *proprio* modo di vedere il mondo. Anche qui, il minimo comune denominatore tra le tecniche algoritmiche in questione è il raggiungimento di scopi predefiniti, su cui non si basa solo la profilazione, ma anche la personalizzazione come mediazione o filtro delle informazioni a noi mostrate come disponibili. Si pensi ai sistemi di raccomandazione che ci suggeriscono quasi ubiquitariamente come dovremmo scegliere, ad esempio, dove andare o quale percorso preferire (pensiamo ai sistemi GPS), che cosa comprare, ascoltare (si pensi a Pandora o Spotify) guardare (si considerino gli algoritmi di raccomandazione da YouTube e Netflix a Amazon Prime Video e Apple Tv, solo per citarne alcuni), ma anche a che ora e che tipo di medicina assumere, quanti passi eseguire al giorno, che cosa e quando mangiare. Esempi di questi ultimi sono gli algoritmi che regolano nuove IoT che connettono bilance digitali e frigoriferi digitali in grado di strutturare la presentazione di opzioni (tangibili) quali alimenti sulla base della tipologia di fisicità corporea e dieta che l'algoritmo ha calcolato e raccomandato di seguire. L'esempio citato è emblematico di come le opzioni informazionali possano concretizzarsi e persino materializzarsi, non solo, come menzionato, nell'accesso a particolari beni, servizi fondamentali e/o opportunità, ma anche in veri e propri oggetti “incarnati” (come degli alimenti). Come anticipato, queste tecniche di personalizzazione sfruttano i profili identitari e comportamentali, inferiti, scoperti o costruiti, per sviluppare raccomandazioni su misura dell'utente al fine di soddisfare obiettivi pre-definiti, sia dai *provider* dei servizi o compagnie tecnologiche in questione, sia da coloro che le sovvenzionano

per accedere e fruire di tali modelli e spazi di valore: in ogni caso, si tratta di obiettivi definiti da *altro* dagli utenti, ovvero etero-definiti, che tendono a coincidere, di sovente, con la massimizzazione degli utili dei soggetti terzi in questione.

Dalla nostra indagine sul ruolo di *governance* algoritmica, si evince che le tecnologie basate su algoritmi hanno, senza ombra di dubbio, un impatto più che abilitante, costringente o limitante, sulla prima condizione *sine qua non* della nostra libertà di scelta, azione e identità. Denominati come *information gatekeepers* (Shapiro 2000) proprio in virtù del loro ruolo di mediazione e selezione delle informazioni a noi accessibili, gli algoritmi non solo influenzano, ma possono costringere a) la nostra disponibilità di opzioni alternative, intese – come sottolineato – in termini informazionali. In base al modo in cui siamo profilati ed etichettati in comunità secondo affinità identitarie-comportamentali, gli algoritmi di filtro che presiedono il flusso informazionale attraverso le ICT digitali, gli OSP e i SNS e, dunque, gli ambienti *online* in cui siamo immersi in modo quasi totalizzante, personalizzano costantemente il nostro ambiente informazionale; in altri termini, decidono sulla nostra disponibilità *onlife* di opzioni alternative, in termini di contenuti informazionali. In primo luogo, gli algoritmi di filtro (quantitativo e qualitativo) *decidono* – secondo pertinenza e profilazione predittiva – oltre che dal punto di vista quantitativo, *quali* informazioni saranno parte dei nostri contesti informazionali; in altre parole, essi “legiferano” su ciò che sarà o meno oggetto di attenzione, elaborazione e analisi da parte dell’utente; dunque, ciò a cui sarà o meno esposto a livello informazionale. Inoltre, tecniche di personalizzazione come gli algoritmi di classificazione decidono e strutturano l’ordine di presentazione di tali opzioni informazionali (si pensi ai risultati filtrati dagli algoritmi alla base dei motori di ricerca Internet quali Google Chrome e Safari o alle *homepage* dei principali SNS o aggregatori di notizie e informazioni). Come affermato in precedenza, tali azioni non sono mai neutrali e solo parzialmente orientate sulla base delle preferenze o degli interessi espressi dall’utente, ad esempio, nell’inserire o interrogare un servizio attraverso una certa *query*. La pertinenza tematica, infatti, è un elemento utilizzato sempre in funzione al raggiungimento dell’obiettivo algoritmico predefinito. Ciò significa che il filtro delle informazioni a noi più rilevanti e la prioritizzazione o strutturazione delle opzioni informazionali assicurano ai *provider* tecnologici che – alla luce della soddisfazione di un certo obiettivo predefinito (come la massimizzazione del click) – alcuni contenuti informazionali (come quelli sponsorizzati) risultino prioritari rispetto ad altri, sebbene parimenti – se non più – rilevanti, ma gratuiti o a sfavore dell’obiettivo

predefinito, così da informare e, dunque, influenzare necessariamente la percezione e la formazione di conoscenza dell'utente, continuando con l'esempio precedente, su di una questione interrogata. Questa tipologia di azione può assumere forme anche più sottili e incisive grazie all'uso di altre tecniche algoritmiche di personalizzazione informazionale quali i sistemi di raccomandazione. Si pensi, ad esempio, alle raccomandazioni vocali (da Siri e Cortana ad Amazon Alexa ed Echo), che ci presentano solo una opzione o poche più in risposta a una particolare *query*, oppure a quei costanti suggerimenti informativi (“hai visto questo, dovresti vedere anche...”) che riceviamo nella forma di annunci pubblicitari (*ads*) o *pop-up* che compaiono interrompendo le pagine web che consultiamo, la musica che ascoltiamo, i video che guardiamo o i social media in uso, raccomandando nella forma di spinte gentili o *nudge* (Sunstein e Thaler 2009; Burr, Cristianini e Ladyman 2018; de Vries 2010) alcuni contenuti informativi piuttosto che altri. Descritti in letteratura come delle “trappole appiccicose”, in virtù della loro capacità di incollare l'utente a specifici contenuti (o opzioni) informativi (Seaver 2018), i sistemi di raccomandazione finiscono per costringere l'utente alla ricezione, visualizzazione e/o considerazione – spesso anche indesiderata – di contenuti o opzioni informativi appositamente selezionati, poiché funzionali alla realizzazione di un obiettivo predefinito. Infatti, queste tecniche sfruttano il profilo predittivo sviluppato sull'utente per bersagliarlo (*targeting*) con raccomandazioni di contenuti o opzioni informativi rispondenti a caratteristiche e preferenze dell'interessato, a loro volta dedotte sia sulla base delle scelte passate (storico) dell'utente (ovvero del suo profilo identitario-comportamentale), attraverso tecniche come il *content-based filtering*, sia sulla base delle scelte effettuate da altri utenti con caratteristiche simili (con profili identitari-comportamentali dedotti come simili), attraverso tecniche di *collaborative filtering*. Le raccomandazioni di tali opzioni, a loro volta, vengono utilizzate per comprendere a che cosa l'utente dà più valore; in altre parole, sulla base del comportamento di risposta dell'utente a una determinata raccomandazione algoritmica, tali tecniche possono scoprire quali tra le caratteristiche inferite ha più presa sull'utente o potenziale di muoverne o innescarne un certo comportamento di risposta: quali sono quegli elementi che, dati certi stimoli informativi, guidano o innescano con maggiore probabilità determinati comportamenti di scelta, contribuendo, in questo modo, a una maggiore personalizzazione del contesto di opzioni e degli stimoli informativi, sempre più cuciti su misura, e – di conseguenza – a perfezionare o affinare la profilazione identitario-comportamentale predittiva sull'utente.

In questo modo, le tecniche in questione non solo influenzano, ma ridefiniscono i contesti di scelta dell'utente, filtrando e ristrutturando il contesto di opzioni informazionali disponibile all'individuo. Questa azione di ridefinizione informazionale assume anche la forma di una pre-determinazione dei contesti o della disponibilità delle opzioni informazionali alternative al soggetto, nella misura in cui sono pre-selezionate algoritmicamente: scelte poiché analizzate e individuate dagli algoritmici come capaci di ri-orientare il comportamento di scelta dell'utente verso determinate opzioni (informazioni, acquisti, contatti, ad esempio) funzionali al raggiungimento di un obiettivo predefinito (ad esempio: aumento dell'attenzione e della presenza nella piattaforma o incremento del click su opzioni quali acquisti, abbonamenti e annunci pubblicitari sovvenzionati). In questo senso, le azioni algoritmiche descritte, decidendo *quali* opzioni informazionali saranno a noi accessibili o disponibili (e sulla base di quali parametri e obiettivi), *costringono* i nostri contesti di scelta, azione e sviluppo identitario: infatti, predeterminando la nostra disponibilità (o *range*) di opzioni informazionali, predeterminano anche quegli elementi che motivano le nostre scelte, azioni e, nel tempo, identità (ciò che può divenire motivo delle mie scelte e dunque informare o orientare lo sviluppo identitario, soprattutto morale). In altre parole, il nostro "comportamento di scelta" (e non solo: si pensi allo sviluppo identitario) può dirsi vincolato all'interno di uno spazio *onlife* di opzioni informazionali – che a loro volta incorporano ciò può informare e motivare o, in altri termini, ciò che possiamo avallare come motivi delle nostre scelte e azione (valori, credenze, opinioni, preferenze, progetti) – algoritmicamente pre-definite come più pertinenti per noi.

Questa prima azione algoritmica sulla nostra disponibilità di opzioni alternative si costituisce come una etero-definizione e predeterminazione della nostra disponibilità informazionale e si rivela costrittiva rispetto alla prima condizione *sine qua non* della nostra libertà di scelta, in quanto mina la condizione di *scegliere e agire altrimenti* da ciò che *de facto* gli algoritmi decidono di mostrarci. In tale senso, tale etero-definizione e predeterminazione algoritmica dei nostri contesti (opzioni) di scelta, della nostra disponibilità o *range* di azioni alternative, solleva il rischio di rendere le nostre scelte algoritmicamente predeterminabili poiché "chiuse" all'interno di uno spazio di opzioni informazionali (relazioni, valori, credenze, preferenze, progetti) definito algoritmicamente, o etero-definito secondo obiettivi che, di sovente, non coincidono con quelli dei soggetti interessati. Questa azione, inoltre, non si limita a una pre-definizione dei contesti di scelta, ma rischia di minare anche l'elemento chiave che caratterizza questa condizione *sine qua non*, ovvero l'elemento di *alternatività* dei corsi di azione

possibili al soggetto alla base della libertà di scelta nel processo decisionale, la quale – sosteniamo – può essere garantita solo da un livello sufficiente di eterogeneità informazionale: solo opzioni che incorporano credenze, opinioni, preferenze, relazioni, progetti, pratiche e valori (e non solo) tra loro diversi possono davvero garantire un'apertura a corsi di azione che siano davvero alternativi. Contesti informativi simili, caratterizzati da un livello ridotto di eterogeneità nelle credenze, nelle opinioni, preferenze, progetti e valori aprono difficilmente a corsi di azioni possibili diversi da quelli che l'individuo è già propenso a seguire. L'azione algoritmica, come anticipato in precedenza, non agisce, infatti, solo sulla pre-selezione di opzioni quali, ad esempio, due possibilità di lavoro diverse tra loro o due destinazioni diverse, come disponibili all'utente. Quando ci rivolgiamo alla questione della libertà nello sviluppo identitario e, nello specifico, alla libertà di sviluppare identità sociali, politiche e, soprattutto, morali autentiche, esito (dinamico) di scelte avallate liberamente nel tempo dall'individuo, è necessario porre in evidenza come una tale azione di predeterminazione algoritmica agisca anche sul portato etico-normativo, oltre che su quello fattuale, socio-politico e culturale delle opzioni informazionali rese disponibili al soggetto. In questo senso, contesti di scelta o ambienti con una ridotta esposizione a preferenze, credenze, valori, scelte, orientamenti, pratiche e caratteristiche diverse o alternative da quelle caratterizzanti un particolare individuo possono comprometterne la libertà di compiere scelte e azioni, oltre a quelle di mero acquisto, anche connotate dal punto di vista sociale, politico e, soprattutto morale, autentiche, ovvero i cui motivi sono stati avallati criticamente dal soggetto – sono queste scelte, infatti, a svolgere un ruolo di rilievo e/o incisivo, nel tempo, nello sviluppo identitario morale. Una esposizione informazionale eterogenea è difatti cruciale affinché l'individuo possa agire davvero altrimenti o in modo alternativo a come *de facto* è nella condizione di fare in un determinato contesto, poiché consente di valutare criticamente se ciò che motiva le relative scelte, come credenze, pensieri, valori, progetti (solo per citare alcuni degli elementi motivanti certe scelte rispetto ad altre e rilevanti per lo sviluppo identitario) è *ottimale* o se invece necessita di essere modificato, rivisto o cambiato e questo è reso possibile dall'esposizione critica a pratiche sociali, culturali, morali e sistemi di preferenze, credenze e valori, orientamenti e modi di fare, inclusi, diversi o alternativi ai propri. In questo senso, un'esposizione informazionale e relazionale eterogenea è un elemento cruciale per uno sviluppo identitario che possa dirsi davvero autentico.

Come evidenziato in precedenza, la personalizzazione dei contesti di scelta informativi basata su tecniche algoritmiche di profilazione quali

il *collaborative filtering* tende a categorizzare gli utenti sulla base di caratteristiche simili, così situandoli in contesti informazionali (e relazionali) costituiti da *like-minded*, da persone “simili”, o meglio, con caratteristiche e/o sistemi di preferenze, credenze e valori inferiti come affini; in altre parole: situandoli in quelle che sono state definite nella letteratura sul tema come *filtering bubbles* (Pariser 2011) o *echo-chambers* (Sunstein 2017): contesti di opzioni (relazioni incluse) simili per persone con profili identitario-comportamentali inferiti o costruiti come simili che tendono, dunque, a configurarsi come delle vere e proprie bolle di filtro in cui l’esposizione informazionale e relazionale è algoritmicamente filtrata – o pre-selezionata – in un modo che tende a ridurre l’eterogeneità di contatti, idee, valori, pratiche, credenze (e non solo), strutturandosi così come delle “casce di risonanza” delle caratteristiche, preferenze, credenze, degli orientamenti e dei valori del singolo e, con esso, del gruppo nel suo insieme. Come argomentato nella letteratura (Parsell 2011; Pariser 2011; Sunstein 2008, 2017), questi ambienti tendono a favorire non solo una “chiusura in sé” dei gruppi in questione, favorendo fenomeni quali la polarizzazione sociale, che minano la possibilità di una sfera pubblica aperta e coesa (Sunstein 2008, 2017; Habermas 2022), ma anche una chiusura in sé, con la conseguente radicalizzazione delle identità degli individui categorizzati e situati in gruppi diversi, in modi che tendono a disabilitare la deliberazione razionale, il dialogo o il confronto critico e la possibilità di cooperazione congiunta (condizioni cruciali anche per il progresso morale) tra gruppi profilati/trattati algoritmicamente come diversi. In questo senso, gli algoritmi creano un vincolo di tipo *epistemologico* sulla nostra libertà di scelta e di azione, con implicazioni particolarmente problematiche per lo sviluppo di identità morali autentiche: essi predeterminano ciò che informa e può motivare le nostre scelte, azioni e, nel tempo, orientare e definire le nostre identità, preselezionando la nostra disponibilità di opzioni alternative secondo obiettivi e fini perlopiù etero-interessati ed etero-definiti rispetto al soggetto, su cui esso può esercitare un controllo minimo, essendogli preclusa la conoscenza sia di come è profilato sia di come risulta categorizzato e, di conseguenza, di quali opzioni (*chances*, opportunità, beni, oggetti, relazioni, ma anche valori, preferenze, credenze, pratiche, progetti ecc.) lo stesso sarà o meno privato, rendendo così sempre meno concreta o accessibile la possibilità di agire altrimenti da che gli è algoritmicamente reso disponibile, presentato con priorità e raccomandato. Inoltre, tale predefinitività informazionale (e relazionale) algoritmica erode la possibilità di sviluppare identità morali autentiche, minando ciò che sta alla base della possibilità di scegliere e agire *davvero* altrimenti: un’esposizione relazio-

nale eterogenea e la possibilità di vagliare criticamente ciò che motiva le nostre scelte e azioni e, attraverso esse, *dà forma alle nostre identità*, favorendo invece lo sviluppo o cristallizzazione di identità sempre più chiuse ed etero-definite. A tal riguardo, si pensi a come ci siano distorsioni nei processi decisionali umani, ormai rese note dalle scienze cognitive, quali, ad esempio, il *confirmation bias*, che tendono a disattivare la riflessione o il pensiero critico dell'individuo nei confronti di chi o che cosa ne conferma il relativo orientamento di scelta o di preferenza iniziale, corroborandolo e persino radicalizzandolo quando situato in contesti di affini.⁵

Abbiamo definito questo vincolo epistemologico poiché tende ad agire in modo diretto sul modo o i contesti in cui formiamo la conoscenza (e i motivi) alla base delle nostre scelte, azioni e identità, e solo in modo indiretto su queste ultime, ovvero a livello pratico *stricto sensu*. Infatti,

5 Rispetto a questa azione algoritmica, come descritta, sulla disponibilità di opzioni informazionali alternative, si potrebbe obiettare che anche le ICT tradizionali, operando certe scelte informazionali e valoriali (ad esempio, selezionando le notizie che sono meritevoli di essere riportate, promuovendo certi prodotti e modelli e non altri, ecc.), possono restringere in modi analoghi alle ICT algoritmiche il *range* delle opzioni disponibili nei nostri contesti di scelta. Sembra plausibile chiedersi, dunque, se l'azione di personalizzazione operata dalle ICT algoritmiche possa minare la libertà di scelta in modo inedito rispetto alle ICT tradizionali. Ciò che rende l'azione delle ICT algoritmiche inedita sulla nostra libertà di scelta rispetto alle ICT tradizionali non è solo la pervasività del mezzo e del suo uso. Su questo, Floridi (2014), per riferirsi alla natura pervasiva e dirompente (ambientale o ri-ontologizzante) del mezzo, scrive che noi non viviamo *nella* radio o televisione ma, al contrario, siamo *immersi* (dentro) gli spazi *online*. Questo distinguo, benché rilevante, non risponde a coloro che affermano come guardare quotidianamente un canale televisivo, connotato da un certo orientamento politico, in modo prolungato nel tempo, potrebbe suscitare effetti simili a quelli di cui sopra (riduzione dell'eterogeneità, chiusura, radicalizzazione, ecc.). Quando si parla di libertà di scelta, come sostenuto per altre questioni etiche dalla O'Neil (2017), è la possibilità di profilazione predittiva a grana fine su base individuale a rendere senza precedenti la personalizzazione informazionale (pre-definizione del *range* di opzioni) operata dalle ICT algoritmiche e, dunque, l'impatto degli algoritmi sulla libertà di scelta. La possibilità di prevedere che cosa può innescare un certo tipo di effetto in una particolare persona per poi deciderne il *range* di opzioni disponibili e/o indirizzarle – tramite tecniche di *micro-targeting* – contenuti scoperti come capaci di indurne specifiche modifiche comportamentali è qualcosa, al momento, operabile solo algoritmicamente. Inoltre, come sostenuto da altri studiosi del settore (Parsell 2008, p. 43; Sunstein 2008, p. 99), sebbene l'eterogeneità informazionale non sia di per sé garantita automaticamente negli spazi analogici, una esclusione dell'inaspettato, nuovo, diverso o divergente dai nostri ambienti informazionali analogici, incarnati o non *online* (se possiamo ancora definirli in questo modo) così sistematica – e su larga scala – è molto difficile da attuare.

tale condizionamento, benché forte, dato lo stato di iperconnessione delle nostre vite e, in particolare, l'enorme quantità di tempo che effettivamente trascorriamo o in cui interagiamo con ICT, OSP e SNS (oltre a tutte le app di IoT), non ci costringe o vincola *stricto sensu* ad agire in un certo modo, o a compiere certe scelte, rispetto ad altre, sebbene, alla luce della riflessione sopra sviluppata, le renda altamente più probabili. Infatti, benché viviamo in uno stato di quasi ubiqua iperconnessione, siamo ancora piuttosto lontani dal vivere spazi completamente ed esclusivamente *online* – quali, ad esempio, potrebbero divenire i cosiddetti metaversi futuri. I contesti incarnati che abitiamo non ci precludono totalmente la possibilità di incontri inaspettati e diversificati e, quindi, di ampliare in modo diverso o eterogeneo la nostra esposizione informazionale e relazionale, azionando così momenti di riflessione critica su ciò che (es. fatti, notizie, esperienze, relazioni, affiliazioni) sta informando il nostro sviluppo di orientamenti, credenze, valori e progetti e sta guidando le nostre scelte e, nel tempo, la formazione delle nostre identità sociali, politiche e morali.

Tuttavia, le tecniche algoritmiche, come attualmente progettate e impiegate, possono dare origine anche a condizionamenti più forti o vincolanti, che agiscono direttamente sulle nostre scelte e azioni, non più solo informandole attraverso la pre-definizione dei contesti epistemologici in cui formiamo e maturiamo i motivi alla base delle nostre scelte e azioni, ma predeterminandole, erodendo l'ultimo baluardo di garanzia minima della nostra libertà di scelta, ovvero la nostra autonomia, quale seconda condizione *sine qua non* della nostra libertà di scegliere, di agire e di sviluppare – nel tempo – identità sociali, politiche e morali autentiche: la possibilità di scegliere e agire secondo motivi di cui l'agente può dirsi in controllo, sviluppando identità morali su cui possiamo rivendicare un grado sufficiente di autorialità (autonomia come principio di autodeterminazione identitaria).

Abbiamo precedentemente visto come i sistemi di raccomandazione siano in grado di scoprire e testare quali elementi – a grana fine e su base individuale – abbiano il potenziale di innescare (*trigger*) determinati comportamenti di risposta. Si pensi, ad esempio, alla facilità con cui gli algoritmi possono scoprire eventi traumatici, vulnerabilità profonde e/o stati di fragilità psicologica ed emotiva, solo per citare alcune informazioni che possono minare la nostra autonomia nel processo decisionale, intesa appunto come controllo sui motivi alla base di certe scelte e azioni; controllo espresso nella possibilità di avallarli criticamente come motivi delle nostre scelte e azioni. Continuando con l'esempio sopra, da una semplice ricerca su Google, dalle app di salute o gestione del ciclo mestruale incorporate nei nostri *devices* o da semplici spostamenti, ad esempio, verso un ospedale

o clinica abortiva, è possibile predire informazioni sensibili, come se una persona stia per vivere o abbia vissuto un aborto (esempi simili possono essere sviluppati rispetto ad altre questioni socialmente e moralmente rilevanti, anche oltre la sfera della salute). Sebbene, nel nostro esempio, si tratti di inferenze di dati di salute, è bene sottolineare come questi ultimi non vengano trattati e/o riutilizzati dagli algoritmi direttamente come tali, quanto piuttosto alla stregua di meri interessi espressi dall'utente nel suo comportamento *onlife* (es. ricerca su Google o spostamento tramite Google map), così eludendo la tutela giuridica introdotta dal GDPR e rendendo vane molte delle attuali tecniche *privacy preserving* e di *data protection* a loro protezione. Tali inferenze costituiscono conoscenza di grande valore economico, e nello specifico, di tipo predittivo-comportamentale (nonostante siano di valore morale neutrale agli algoritmi), poiché forniscono un criterio per la scelta di quali contenuti informativi pre-selezionare e scegliere per bersagliare l'individuo al fine di ottenere comportamenti di risposta allineati a promozione di obiettivi pre-determinati. Ad esempio, da alcune mie ricerche su Google (esempio: sui sintomi del cancro) e da una mia passata geo-localizzazione (esempio: in una struttura sanitaria), e la correlazione di tali dati ad altri di altri utenti presentanti caratteristiche simili, è possibile dedurre, inferire, scoprire, etichettarmi, trattarmi come malata e/o interessata ad alcune informazioni concernenti tale patologia. Alla luce di questa inferenza, i sistemi di raccomandazione, al fine di massimizzare il loro obiettivo predefinito, come, ad esempio, la massimizzazione del *click-through rate*, potrebbero iniziare a iper-consigliarmi (*hyper-nudge*) o iper-spingermi (*push*) ubiquitariamente, ovvero in tutti i miei dispositivi interconnessi, opzioni informative pre-selezionate, sia in quanto pertinenti alla mia ricerca, sia in quanto popolari, ovvero scelti da molte persone con caratteristiche simili (es. *collaborative filtering*), sia e soprattutto poiché sponsorizzate da compagnie farmaceutiche, assicurativo-sanitarie, o strutture ospedaliere private, fino a fondazioni di ricerca sperimentale o associazioni contro le cure mediche. Tali elementi informativi assumono la forma di *ads* o suggerimenti di contatti che compaiono ben oltre, ad esempio, una singola richiesta in un OSP, interrompendo la visualizzazione di contenuti o di qualsiasi altra attività (guida, ascolto di un podcast, visione di un video, per citarne alcune). Questa tipologia di raccomandazione, sulla base del contenuto informativo scelto per innescare un certo comportamento di scelta verso un obiettivo predefinito, non si configura più come una spinta gentile (*nudge*), ma persistente (*push*) – si pensi, ad esempio, a raccomandazioni di contenuti informativi quali immagini forti (un malato durante terapia medica), informazioni partico-

larmente sensibili (ad esempio sulla rinuncia della cura medica) –, ovvero capace di innescare un comportamento di risposta tale da determinare in modo cruciale il mio comportamento di scelta e azione (ad esempio, rinuncia al trattamento medico), by-passando la fase di vaglio o avallo critico di ciò che sta muovendo la mia scelta verso una direzione o corso di azione specifico rispetto a un altro. Queste raccomandazioni o spinte persistenti, dal portato informazionale e di innesco forte, possono minare la nostra autonomia sia come controllo sia come avallo critico, quale elemento etico-normativo che connota il nostro agire con una dimensione pratico-morale. In altre parole: tali raccomandazioni o spinte agiscono sui motivi alla base di determinate scelte e azioni, trasformando l'opzione informazionale (e ciò che incorpora) raccomandata – opzione preselezionata in modo eteronomo, ovvero dagli algoritmi, come cucita su mia misura, ovvero su ciò che ha presa e capacità di innescare risposte comportamentali automatiche – dall'essere potenziale *motivo*, che posso avallare, all'essere *causa* strettamente determinante le mie scelte, azioni e, nel tempo, la mia identità; in altri termini: da informare (consigliare/raccomandare) a determinare (decidere) le nostre scelte, azioni e, nel tempo, le nostre identità, al nostro posto.

Questa tipologia di azione algoritmica crea un condizionamento o vincolo forte sulla nostra libertà di scegliere, agire e, nel tempo, di sviluppare identità morali autentiche, che va oltre il vincolo di natura epistemologica definito sopra, quale azione sulla prima condizione *sine qua non* della libertà di scelta, ovvero sulla nostra possibilità di agire altrimenti. Le tecniche algoritmiche menzionate, infatti, non solo possono vincolare i potenziali motivi alla base delle nostre scelte (conoscenza, esperienze, credenze, affiliazioni, valori, preferenze, progetti ecc.) all'interno di uno spazio di opzioni etero-definito o pre-definito/predeterminato algoritmicamente, ma possono anche sfruttare tale pre-selezione di opzioni già personalizzate sui profili degli utenti per testare quali elementi informativi raccomandare (o spingere) al fine di ottenere comportamenti di scelta allineati alla realizzazione di determinati obiettivi, a discapito delle conseguenze sull'individuo a breve (azioni) e a lungo (identità) termine. Questo secondo vincolo è di tipo *pratico*, proprio poiché agisce direttamente sul modo in cui compiamo scelte e azioni, sospendendo l'esercizio dell'elemento normativo che caratterizza il nostro agire come morale o pratico, ovvero dell'avallo riflessivo o critico (*reflective endorsement*) – su cui si fonda la nozione di controllo come autogoverno nel processo decisionale – dei motivi alla base di certe scelte, azioni e comportamenti. In altre parole: esso va a erodere la seconda condizione *sine qua non* all'esercizio della nostra libertà di scegliere, agire e sviluppare identità morali autentiche, ovvero mina la nostra

autonomia, e così, la possibilità di autorialità e, dunque, di autenticità nello sviluppo identitario, innescando l'attuazione di scelte e azioni a-riflessive, dettate *da altro* dal soggetto, ovvero mosse da obiettivi e fini a breve termine ed etero-definiti, disinteressati agli interessati, i quali, invece, divengono, per richiamare una terminologia di stampo kantiano, meri mezzi per il soddisfacimento di obiettivi preselezionati da algoritmi.

3. Criteri di design etico: intelligibilità, eterogeneità e confidenzialità informazionale

Nella sezione precedente abbiamo argomentato come l'attuale combinazione di tecniche algoritmiche di profilazione e personalizzazione che governano *by design* le ICT che, in modo ubiquo, permeano i nostri ambienti informazionali stiano gradualmente inficiando le condizioni *sine qua non* alla base della nostra libertà di scelta e azione *onlife* e, con esse, erodendo la possibilità sviluppare, nel tempo, identità, non solo socio-politiche, ma anche morali, che possano dirsi autentiche e di cui possiamo dirci gli autori. In particolare, abbiamo illuminato come tali sistemi stiano dando o possano dare origine a vincoli invisibili e intangibili sia di carattere epistemologico sia di carattere pratico *stricto sensu* sul nostro "spazio" di libertà di scelta e di azione, con implicazioni etiche allarmanti che intaccano sia lo sviluppo identitario individuale, ma anche lo sviluppo di identità di gruppo e la possibilità di dialogo e cooperazione tra gruppi con identità diverse, aprendo a riflessioni ulteriori sulle ricadute critiche di questi fenomeni per la creazione di una sfera pubblica coesa e la possibilità di progresso morale su larga scala. Abbiamo anche mostrato come le attuali soluzioni avanzate nell'ambito della *privacy* e della *data protection* non siano, da sole, sufficienti alla tutela della libertà di identità e, nello specifico, alla libertà di sviluppare identità morali autentiche negli ambienti datificati e iper-profilati delle nostre società dell'informazione, proprio a causa della pervasiva penetrazione e *governance* degli algoritmi esercitata oggi a livello informazionale. Tali soluzioni e riflessioni, infatti, non tengono conto in modo sufficiente o sufficientemente adeguato delle dimensioni essenziali (o *sine qua non*) all'esercizio della libertà di identità come profondamente interconnessa alla libertà di scelta e di azione dell'individuo: condizioni imprescindibili, come evidenziato nella sezione precedente, per la tutela di caratteristiche quali l'apertura, l'autorialità e l'autenticità dell'individuo nello sviluppo identitario sia socio-politico e, soprattutto, morale. In questa ultima sezione, mostriamo, in breve, come l'indagine etica condotta sull'impatto de-

gli algoritmi sulle condizioni *sine qua non* della libertà di scelta, azione e di identità (soprattutto: morale autentica) *onlife* permetta di fare luce e indicare alcune condizioni (o criteri) che richiedono di essere considerati e, soprattutto, operazionalizzati *by design* nello sviluppo di sistemi o ICT algoritmiche, che si dimostrino interessate alla tutela dell'identità individuale negli ambienti iper-connessi e iper-profilati contemporanei, invitando la ricerca futura alla loro attenzione, rielaborazione e/o, su di essi, ulteriore lavoro sia concettuale sia tecnico alla luce della questione in oggetto.

L'analisi etica dell'impatto degli algoritmi sulla prima condizione *sine qua non* della libertà di scelta, ovvero la possibilità di fare altrimenti, o disponibilità di opzioni alternative, ha illuminato come gli algoritmi – e, nello specifico, tecniche algoritmiche di personalizzazione e profilazione – nel ruolo di nuovi *gatekeepers* dell'informazione stiano sollevando un'azione vincolante la nostra libertà di scelta, azione e identità che assume la forma di una etero-definizione o predeterminazione del contesto di opzioni informazionali disponibili al soggetto nel processo decisionale. In sintesi: tale etero-definizione o predeterminazione algoritmica agisce pre-definendo ciò che andrà a informare e/o motivare – e/o costituire in sé opzioni avallabili – il processo decisionale dell'individuo, vincolandone così il comportamento di scelta all'interno di uno spazio di opzioni predefinito dagli algoritmi. Questa azione di pre-definizione si basa sulla scoperta o costruzione di un profilo predittivo dell'individuo (identitario-comportamentale) assegnato e poi sfruttato per categorizzarlo in gruppi di simili a cui indirizzare, presentare, raccomandare contenuti o contesti informazionali specifici, pre-selezionati in quanto funzionali al soddisfacimento di obiettivi etero-definiti e etero-interessati, perlopiù disallineati rispetto a quelli del soggetto, compromettendone così la possibilità di fare altrimenti da ciò che gli è algoritmicamente presentato. Al fine di contrastare o prevenire questa azione di etero-definizione o pre-definizione algoritmica e, dunque, di salvaguardare – a una soglia minima – la disponibilità di opzioni alternative quale prima condizione *sine qua non* della nostra libertà di scelta, azione e identità, la progettazione di tali tecnologie dovrebbe rispondere a un *criterio di intelligibilità by design*⁶ che garantisca la rivelazione (o

6 Questo criterio etico, che proponiamo in relazione alla tutela delle condizioni della libertà di scelta, risponde – enuclea maggiormente, contestualizza e traduce nella pratica – a uno dei principi etici chiave messi in luce dal gruppo di esperti di alto livello sull'IA selezionato dall'UE negli *Orientamenti etici per un'IA affidabile* (2019); documento oggi cardine nella letteratura sulla progettazione etica dei sistemi basati su algoritmi: il principio di esplicabilità, che richiede il *design* di sistemi che operino in modo trasparente o almeno spiegabile ai soggetti

disclosure by design) all'individuo e la relativa comprensione (attraverso la scelta di linguaggi sensibili all'utente in questione) del modo in cui è stato profilato e dunque categorizzato, e sulla base di quali ragioni, ovvero dati e informazioni inferiti sul soggetto interessato, così che lo stesso possa rivendicare non solo il diritto di conoscere o sviluppare consapevolezza su ciò su cui si basa la selezione delle opzioni informazionali disponibili, ma e soprattutto il diritto di rivedere e modificare o sospendere l'uso dei parametri utilizzati. In questo modo, l'individuo può riacquisire un grado minimamente sufficiente di controllo sia sul profilo-identitario comportamentale predittivo assegnatogli, prima definito esclusivamente secondo parametri scelti algoritmicamente, sia su quei contenuti o opzioni informazionali a esso accessibili, che dunque vanno a informare inevitabilmente ciò che può orientare o che va a formare il relativo comportamento di scelta (controllo e presenza del soggetto sull'azione di predeterminazione algoritmica).

La tutela della possibilità di fare altrimenti, inoltre, richiede anche che la categorizzazione dell'individuo in contesti o spazi informazionali filtrati e pre-selezionati algoritmicamente mantenga sempre un grado minimamente sufficiente a garantire corsi di azione davvero alternativi, ovvero, una esposizione informazionale attraverso una disponibilità di opzioni informazionali che incorporano un livello sufficiente di eterogeneità informazionale e relazionale, tale che possa garantire sia apertura, sia, e soprattutto, spronare il vaglio critico/riflessivo su ciò che può motivare scelte e azioni. La tutela della prima condizione *sine qua non* della libertà di scelta, di azione e di identità richiede dunque di rispondere a un criterio sia di a) intellegibilità sia di b) *eterogeneità* informazionale, per tutelare la nostra libertà di identità e, nello specifico, di sviluppo identitario morale autentico. È possibile trovare alcuni primi esempi di operazionalizzazione dei criteri menzionati – sebbene non siano stati sviluppati in riferimento alla questione qui analizzata né tantomeno siano stati concettualizzati nei termini di cui sopra – in un esperimento condotto dal MIT Center for Civic Media del Massachusetts Institute of Technology, denominato Gobo: una piattaforma di SNS e di aggregazione di notizie progettata per porre l'individuo “in controllo”, fornendogli conoscenza sul modo in cui è profilato e categorizzato (sulla

interessati. La descrizione di questo principio etico di alto livello è, per sua natura, generale; l'UE invita, perciò, sia a una sua declinazione o contestualizzazione nei diversi ambiti in cui tali sistemi vengono utilizzati, ovvero in relazione alle diverse e specifiche questioni morali che tali sistemi sollevano; sia allo sviluppo di misure o criteri di *design* per la sua operazionalizzazione. Nella formulazione del criterio di intellegibilità – di cui sopra –, abbiamo cercato di accogliere e sviluppare entrambe le istanze.

base di quali informazioni e regole determinate algoritmicamente). Gobo nasce per testimoniare la realizzabilità o operazionalizzabilità dei criteri in questione, la possibilità tecnica di sviluppare ICT algoritmiche secondo criteri di *ethics by design* che pongono al centro la tutela dell'individuo.

Se i criteri di intelligibilità ed eterogeneità sembrano tutelare, almeno una soglia minima, la possibilità della prima condizione *sine qua non* della libertà di scelta, azione e identità *onlife*, un altro criterio, che definiamo *criterio di confidenzialità*, sembra necessario al fine di preservare a una soglia minima la nostra autonomia, come controllo sui motivi alla base delle nostre scelte e azioni (principio di autodeterminazione identitaria), quale seconda condizione *sine qua non* per l'esercizio della nostra libertà di scelta, azione e identità *onlife*. Abbiamo argomentato come gli algoritmi, o meglio, alcune tecniche combinate di profilazione predittiva e personalizzazione informazionale a grana fine come i sistemi di raccomandazione, possano intaccare l'autonomia come autodeterminazione dell'individuo, bersagliandolo con contenuti (o opzioni) informativi particolarmente sensibili (esempi ne sono i traumi, eventi intimi, stati di salute e/o vulnerabilità profonde), scoperti e poi testati sull'individuo come capaci di innescare il relativo comportamento (di scelta e azione) in modo definito, perlopiù a-riflessivo o istintivo, predicibile e dunque determinabile in anticipo. Abbiamo mostrato come l'uso di tali tecniche e informazioni (o inferenze) sensibili per soddisfare e orientare il comportamento di scelta e azione individuale verso obiettivi etero-definiti e pre-selezionati algoritmicamente sfida la nostra possibilità e capacità di controllo (e avallo critico) dei motivi alla base di certe scelte e azioni, con conseguenze problematiche, soprattutto quando tali scelte si estendono oltre la sfera commerciale e si riversano su contenuti di carattere politico, sociale e morale, finendo per mettere in discussione la nostra reale autorialità sulle nostre scelte e azioni e portando allo sviluppo di identità sempre meno autentiche. Al fine di salvaguardare, almeno a una soglia minima, la condizione di autonomia come controllo (o almeno partecipazione) dell'individuo su ciò che muove le relative scelte è necessario progettare ICT algoritmiche che non prevedano *by design* l'utilizzo di conoscenza (ovvero: inferenze, modelli, informazioni) scoperte come di valore, poiché dotate di contenuto sensibile (con capacità di innesco o spinta comportamentale) per la personalizzazione o per la raccomandazione informazionale volta al soddisfacimento di obiettivi definiti da terze parti. La tutela dell'autonomia richiede dunque a una soglia minima di operazionalizzare il criterio di confidenzialità informazionale, il quale consiste nel progettare ICT algoritmiche che *by design* non sfruttino conoscenza e inferenze sensibili per la realizzazione di obiettivi pre-definiti

algoritmicamente e orientati da interessi di terzi. Se, infatti, da un lato, il comportamento algoritmico in *input*, ovvero, la processazione del reale (con profilazione individuale inclusa) e lo sviluppo di deduzioni, modelli e inferenze predittive su questo materiale non sembra essere soggetto a regolamentazione, a causa della pervasiva presenza algoritmica nelle nostre società informazionali e, dunque, del continuo processo di datificazione a cui il reale è soggetto, dall'altro lato, lo stesso non può dirsi rispetto alla regolamentazione del comportamento algoritmico in *output*, ovvero di che cosa gli algoritmi possono utilizzare (nell'ambito delle inferenze, dei *patterns* e delle correlazioni sviluppate) al fine di soddisfare compiti e obiettivi definiti. La regolamentazione del comportamento algoritmico in *output* può essere effettuata dalla fase di progettazione delle ICT algoritmiche attraverso il criterio di confidenzialità *by design*, che appunto va a regolamentare ciò che, sebbene inferito come di valore economico, può o meno essere utilizzato per obiettivi etero-definiti. L'operationalizzazione di questo criterio etico consiste di due fasi. La prima fase consiste nell'allenare i sistemi in questione al riconoscimento (attraverso apprendimento supervisionato) di quali informazioni inferite potrebbero incorporare contenuti (conoscenza, eventi, caratteristiche, ecc.) di carattere sensibile all'individuo, ad esempio, poiché relative, anche indirettamente, alla sfera della salute psicologica o fisica, o poiché particolarmente riferiti a eventi in genere ritenuti traumatici (solo per citare alcuni esempi). Abbiamo visto, infatti, che di sovente tali informazioni non sono tutelabili attraverso GDPR e sforzi simili, spesso poiché non trattano il dato salute in modo diretto – esempio: non viene sviluppata un'inferenza diretta sul dato di salute “depressione”, ma viene sviluppato o prodotto un profilo, o profilo di gruppo, ovvero una categorizzazione, attraverso l'inferenza “gruppo di persone tristi”, esito dell'associazione tra tempo di presenza degli interessati nei SNS e la loro visione di determinati contenuti *online*, che tuttavia in nessuno dei suoi termini include il riferimento alla patologia. Una volta identificate tali inferenze o informazioni di carattere sensibile sull'individuo, per il principio di intelligibilità, l'utente dovrebbe essere informato, sia per verificarne la correttezza e richiederne la modifica, sia e soprattutto per rivendicarne, qualora desiderata, la tutela e, dunque, il non utilizzo, applicando su di esse tecniche di confidenzialità (o protezione) algoritmica. I *designer* di tali sistemi al fine di essere conformi al rispetto dell'autonomia come condizione *sine qua non* della libertà di scelta, azione e identità, almeno a una soglia minima, dovrebbero predisporre spazi appositi dove sottoporre la conoscenza inferita e predittiva sviluppata al vaglio del singolo interessato e sviluppare tecniche specifiche per garantire la tutela – se rivendicata – di queste infor-

mazioni dalla loro divulgazione, uso e/o confluenza al di fuori dello spazio di loro conservazione, così come – se desiderata – richiederne l’eliminazione. Questo criterio promuoverebbe l’esercizio dell’autonomia dell’individuo su ciò che potrebbe minarne il controllo sulle proprie scelte e azioni e, nel tempo, l’autorialità e autenticità nello sviluppo identitario, innescando comportamenti principalmente a-riflessivi, istintivi e/o dettati da contenuti emotivamente carichi o di carattere sensibile, sfruttati algoritmicamente per soddisfare obiettivi terzi all’interessato.

In quest’ultima sezione abbiamo segnalato alcuni criteri chiave, intelligibilità, eterogeneità e confidenzialità, al fine di tutelare – almeno a una soglia minima – le condizioni *sine qua non* alla base della possibilità di scegliere, agire e, nel tempo, sviluppare identità morali autentiche. Sebbene siamo consapevoli del molto lavoro, sia concettuale sia tecnico, ancora da sviluppare sul tema, ci auguriamo che l’indagine etica e i criteri etici sopra illuminati possano costituire strumenti di progettazione etica soprattutto per i *designer* delle ICT algoritmiche che siano a tutela della nostra libertà di scelta, azione e identità *onlife*, ma anche strumenti di valutazione e regolamentazione etica per coloro (ad esempio, enti istituzionali o fornitori privati) che sono chiamati a una loro adozione e implementazione, anche in ambiti sociali sensibili, affinché siano orientati da un approccio etico che pone le persone al centro, rispettandone la possibilità di libertà, di autenticità e la relativa unicità.

4. Conclusioni

In questo articolo abbiamo cercato di proporre un percorso concettuale alternativo a quello proposto nell’ambito dell’etica dell’IA per porre in luce la relazione tra ICT algoritmiche e identità individuale, con un *focus* particolare sul tema della libertà di sviluppo di identità morali autentiche, su cui il dibattito attuale non si è soffermato finora a sufficienza. Nello specifico, abbiamo mostrato come sia possibile offrire una concettualizzazione sul tema complementare o a integrazione di quella esistente, ma anche più ricca e fruttuosa dal punto di vista filosofico, che riconosce l’importanza di considerare l’identità individuale in termini informativi nelle nostre società iper-connesse odierne, ma supera l’esclusiva concettualizzazione delle sfide all’identità individuale in termini di violazioni della *privacy*, riducendo le soluzioni a sua tutela a tecniche o strumenti legali di protezione dei dati. In particolare, abbiamo mostrato come un’indagine etica sull’impatto delle tecniche algoritmiche sulla nostra libertà di scelta e azione e,

nello specifico, sulle condizioni che ne consentono l'esercizio almeno a una soglia minima, possa illuminare nuove sfide e dimensioni della nostra libertà di identità, che necessitano di essere affrontate e tutelate al fine di salvaguardare la nostra libertà di scegliere, agire e sviluppare, nel tempo, identità morali autentiche nelle nostre società informazionali contemporanee. Nei limiti dell'indagine condotta, speriamo che la messa in luce di tali sfide sulle dimensioni o condizioni cruciali della nostra libertà di scelta, di azione e di sviluppo di identità morali autentiche possa fornire una prima bussola sia di valutazione sia di progettazione etica di tecnologie algoritmiche a tutela della nostra libertà di identità, della nostra autorialità e unicità come individui nell'era dell'iper-profilazione algoritmica.

Riferimenti bibliografici

Albritton, A.

1985 *Freedom of Will and Freedom of Action*, in "Proceedings and Addresses of the American Philosophical Association", 59, 2, pp. 239-251.

Bakardjieva, M., Gaden, G.

2012 *Web 2.0 technologies of the self*, in "Philosophy & Technology", 25, pp. 399-413.

Blume, P.

2002 *Protection of Informational Privacy*, Djøf Forlag, Copenhagen.

Burr, C., Cristianini, N., Ladyman, J.

2018 *An Analysis of the Interaction Between Intelligent Software Agents and Human Users*, in "Minds and Machines", 28, 4, pp. 735-774.

Calo, M.R.

2014 *Digital Market Manipulation*, in "The George Washington Law Review", 82, 4, pp. 996-1012.

De Caro, M.

2004 *Il libero arbitrio. Una Introduzione*, Laterza Editori, Roma-Bari.

De Vries, K.

2010 *Identity, profiling algorithms and a world of ambient intelligence*, in "Ethics and Information Technology", 12, 1, pp. 71-85.

Ess, C.

2010 *The embodied self in a digital age: Possibilities, risks, and prospects for a pluralistic (democratic/liberal) future?*, in “Nordicom Review”, 31, pp. 105-118.

Floridi, L.

2005 *The Ontological Interpretation of Informational Privacy*, in “Ethics and Information Technology”, 7, pp. 185-200.

2006 *Four challenges for a theory of informational privacy*, in “Ethics and Information Technology”, 8, 3, pp. 109-119.

2014 *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford University Press, Oxford.

2016 *On human dignity as a foundation for the right to privacy*, in “Philosophy and Technology”, 29, 4, pp. 307-312.

Floridi, L., Taddeo, M.

2018 *How AI Can Be a Force for Good*, in “Science”, 361, 6404, pp. 751-753.

Frischmann, B., Selinger, E.

2018 *Re-Engineering Humanity*, Cambridge University Press, Cambridge.

Habermas, J.

2022 *Ein neuer Strukturwandel der Öffentlichkeit und die deliberative Politik*, Suhrkamp Verlag, Berlin.

Hobbes, T.

1654 [1999] *Of Liberty and Necessity*, in V. Chappell (a cura di), *Hobbes and Bramhall on Liberty and Necessity*, Cambridge University Press, Cambridge, pp. 15-42.

Hull, G.

2015 *Successful Failure: What Foucault Can Teach Us About Privacy Self-Management in a World of Facebook and Big Data*, in “Ethics and Information Technology”, 17, 2, pp. 89-101.

Kant, I.

2019 [1787] *Critica della ragione pura*, tr. it. G. Gentile, G. Lombardo-Radice, Laterza Editori, Roma-Bari.

2019 [1788] *Critica della ragione pratica*, tr. it. F. Capra, Laterza Editori, Roma-Bari.

Liao, S.M.

2020 *The Moral Status and Rights of Artificial Intelligence*, in S. M. Liao (a cura di), *Ethics of Artificial Intelligence*, Oxford University Press, New York, pp. 480-503.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L.

2016 *The Ethics of Algorithms: Mapping the Debate*, in “Big Data & Society”, 3, 2, pp. 1-21.

Mori, M.

2001 *Libertà, necessità, determinismo*, Il Mulino, Bologna.

Nissenbaum, H.

2010 *Privacy in context: technology, policy and the integrity of the social life*, Stanford University Press, Stanford.

O’Neil, C.

2017 *Armi di distruzione matematica. Come i big data aumentano la disuguaglianza e minacciano la democrazia*, Bompiani, Milano.

Pariser, E.

2011 *The Filter Bubble: What the Internet is Hiding from You*, Viking, London.

Parsell, M.

2008 *Pernicious Virtual Communities: Identity, Polarisation and the Web 2.0*, in “Ethics and Information Technology”, 10, 1, pp. 41-56.

Pasquale, F.

2015 *The Black Box Society: The Secret Algorithms that Control Money and Information*, Harvard University Press, Cambridge.

Richards, N.M.

2013 *The Dangers of Surveillance*, in “Harvard Law Review”, 126, 7, pp. 1934-1965.

Rosenberg, M.

2018 *Bolton Was Early Beneficiary of Cambridge Analytica’s Facebook Data*, <https://www.nytimes.com/2018/03/23/us/politics/bolton-cambridge-analyticas-facebook-data.html>

Seaver, N.

2018 *Captivating algorithms: Recommender systems as traps*, in “Journal of Material Culture”, 24, 4, pp. 421-436.

Shapiro, S.

2020 *Algorithmic Television in the Age of Large-scale Customization*, in “Television & New Media”, 21, 6, pp. 658-663.

Skinner, Q.

2012 *Liberty before Liberalism*, Cambridge University Press, Cambridge.

Sunstein, C.

2008 *Democracy and the Internet*, in J. van den Hoven, J. Weckert (a cura di), *Information Technology and Moral Philosophy*, Cambridge University Press, Cambridge, pp. 93-110.

2016 *The ethics of influence*, Cambridge University Press, Cambridge.

2017 *#Republic: Divided Democracy in the Age of Social Media*, Princeton University Press Princeton.

Susser, D., Roessler, B., Nissenbaum, H.

2019 *Technology, autonomy, and manipulation*, in “Internet Policy Review”, 8, 2, pp. 1-22.

Steeves, V., Lucock, C., Keer, I. (a cura di)

2009 *Privacy, Identity and Anonymity in a Network World: Lessons from the Identity Trail*, Oxford University Press, New York.

Tene, O., Polonetsky, J.

2013 *Big Data for all: Privacy and user control in the age of analytics*, in “Northwestern Journal of Technology and Intellectual Property”, 11, 5, pp. 240-272.

Thaler, R., Sunstein, C.

2009 *Nudge: Improving decisions about health, wealth and happiness*, Penguin, London.

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., Floridi, L.

2022 *The Ethics of Algorithms: Key Problems and Solutions*, in “AI & Society”, 37, pp. 215-230.

Turow, J.

2011 *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*, Yale University Press, New Haven.

Wachter, S.

2020 *Affinity Profiling and Discrimination by Association in Online Behavioral Advertising*, in “Berkeley Technology Law Journal”, 35, 2, pp. 369-430.

Wachter, S., Mittelstadt, B.

2019 *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, in “Columbia Business Law Review”, 2, pp. 494-620.

Williams, J.

2018 *Stand out of our Light: Freedom and Resistance in the Attention Economy*, Cambridge University Press, Cambridge.

Yeung, K.

2017 *Hyper-nudge: Big data as a mode of regulation by design*, in “Information, Communication and Society”, 20, 1, pp. 118-136.

Zuboff, S.

2019 *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Public Affairs, New York.