

## Proposta di Insegnamento

Corso offerto per la scelta libera a:

Laurea Triennale Ingegneria Informatica (it)

II semestre AA 2021/22. 48 h – 24 slot da 2 h

Crediti 6 cfu

### **Saperi di genere ed etica nell'Intelligenza Artificiale:**

Con il diffondersi delle applicazioni che utilizzano tecniche di Intelligenza Artificiale, sempre maggiore attenzione stanno avendo le possibili implicazioni che tali applicazioni hanno a livello etico. Tale dibattito coinvolge, oltre all'opinione pubblica, le istituzioni politiche, le aziende fornitrici di software e servizi informatici, programmatori e il mondo accademico. Numerose sono le iniziative che promuovono una riflessione a livello normativo e culturale, a titolo di esempio si può citare: lo studio "Orientamenti etici per un'IA affidabile" (2019) promosso dalla Commissione Europea; lo studio "The ethics of artificial intelligence: Issues and initiatives" (2020) promosso dal Parlamento Europeo; il "Preliminary study on the Ethics of Artificial Intelligence" dell'UNESCO (2019). Anche dal punto di vista della formazione, corsi sull'etica dell'IA sono già nati in molte Università, tra cui MIT<sup>1</sup> (promosso dal Media Lab), Stanford<sup>2</sup> (all'interno del Symbolic system program), Columbia University<sup>3</sup> (Computer science), Università di Bologna<sup>4</sup> (LM in Artificial Intelligence).

Si ritiene pertanto utile per i nostri studenti e strategico per un Dipartimento e un Ateneo particolarmente sensibili a questioni etiche come l'inclusione e i saperi di genere promuovere l'istituzione di un corso dal titolo "Saperi di genere ed etica nell'Intelligenza Artificiale", all'interno della Laurea Triennale in Ingegneria Informatica e, in ragione della particolarità delle tematiche trattate, offrendone l'accesso anche agli studenti della Laurea Magistrale.

Al fine di sviluppare un'Intelligenza Artificiale inclusiva ed etica, è opportuno affrontare alcune tematiche che emergono da un'analisi dal punto di vista di genere, di etnia e di sviluppo personale e sociale degli algoritmi di Intelligenza Artificiale. A tal fine è necessario introdurre alcuni concetti di base riguardanti i saperi di genere con particolare attenzione agli stereotipi e ai pregiudizi che condizionano le interazioni sociali e che potrebbero influenzare lo sviluppo di tali algoritmi. Lo studio di metodologie e di tecniche volte alla individuazione e al trattamento di tali possibili bias sussunti in Sistemi basati sull'Intelligenza Artificiale è un obiettivo centrale del Corso. Inoltre, risulta altresì importante fornire conoscenze e principi di base per trattare aspetti etici, giuridici ed economici dell'Intelligenza Artificiale.

### **Promotori**

Silvana Badaloni, Studiosa Senior dello Studium Patavinum, Professoressa associata di Intelligenza Artificiale – fuori ruolo - presso l'Università di Padova.

Antonio Rodà, Ricercatore universitario, Professore aggregato di Fondamenti di informatica, membro del comitato scientifico del General Course in Diritti umani e inclusione dell'Università di Padova.

---

<sup>1</sup> <https://dam-prod.media.mit.edu/x/2018/07/30/Syllabus%20Ethics%20and%20Governance%20of%20AI%20.pdf>

<sup>2</sup> <https://web.stanford.edu/class/cs122/>

<sup>3</sup> <http://www.cs.columbia.edu/~dechant/aisafety18/index.html>

<sup>4</sup> <https://www.unibo.it/it/didattica/insegnamenti/insegnamento/2020/446601>

## **Conoscenze ed abilità da acquisire**

Le studentesse e gli studenti al termine del corso saranno in grado di:

- Comprendere i principali aspetti etici collegati alla creazione e all'uso di sistemi di Intelligenza Artificiale.
- Avere consapevolezza dei saperi di genere e delle tematiche legate all'inclusione.
- Analizzare e discutere problemi etici dovuti all'applicazione dell'IA in alcuni casi di studio notevoli.
- Analizzare e discutere i potenziali bias (di genere, etnia, ecc.) dei sistemi di IA.
- Individuare soluzioni computazionali e metodologiche per attenuare eventuali ripercussioni etiche negative derivanti dall'applicazione di tecniche di IA.

**Prerequisiti** Nessuno

## **Contenuti**

La struttura e le tematiche dell'insegnamento saranno le seguenti:

- Richiami di Intelligenza Artificiale. Ragionamento simbolico e Machine Learning
- Cultura della parità e Saperi di Genere
- Innovazione di genere e inclusione
- Bias negli Algoritmi di Intelligenza Artificiale e metodi di mitigazione
- Robotica, etica e genere
- IA e creatività artistica
- Aspetti etici, giuridici ed economici dell'IA

Il Corso è conformato in Lezioni e Seminari tenuti da Professori/esse invitati da altre Università.

L'insegnamento rientra nei SSD ING-INF/05 e INF/01 con un impianto multidisciplinare che prevede il coinvolgimento di docenti con expertise diverse.

Si prevede lo sviluppo di Tesine di approfondimento da parte degli studenti.

## **Bibliografia**

European Commission (2019). Orientamenti etici per un'IA affidabile.  
[https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=60430](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60430)

UNESCO (2019). Preliminary study on the Ethics of Artificial Intelligence.  
<https://unesdoc.unesco.org/ark:/48223/pf0000367823>

European Parliamentary Research Service (2020). The ethics of Artificial Intelligence: Issues and initiatives.  
[https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS\\_STU\(2020\)634452\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf)

European Commission (2020). Gendered Innovation 2: How Inclusive Analysis Contributes to Research and Innovation ISBN 978-92-76-16416-62.  
[https://ec.europa.eu/info/news/gendered-innovations-2-2020-nov-24\\_en](https://ec.europa.eu/info/news/gendered-innovations-2-2020-nov-24_en)

R. Chatila, F. Rossi (2021). AI fairness is an economic and social imperative. Here's how to address it. The Davos Agenda. <https://www.weforum.org/agenda/2021/01/how-to-address-artificial-intelligence-fairness/>

Stuart Russell, Peter Norvig, Artificial Intelligence: A modern approach, Third Edition. Prentice Hall, 2010.

N. Bostrom and E. Yudkowsky. 'The ethics of artificial intelligence'. In W. M. Ramsey and K. Frankish, editors, The Cambridge Handbook of Artificial Intelligence, pages 316–334. Cambridge University Press, Cambridge, 2014.

S. J. Russell, D. Dewey, and M. Tegmark, 'Research priorities for robust and beneficial artificial intelligence', *AI Magazine*, 2015

Armstrong, S., Sandberg, A. & Bostrom, N. (2012). Thinking Inside the Box: Controlling and Using an Oracle AI. *Minds & Machines* (2012) 22: 299–324.

N. Soares, 'The Value Learning Problem'. Ethics for Artificial Intelligence Workshop at 25th International Joint Conference on Artificial Intelligence (IJCAI-2016) New York, NY, USA 9–15 July 2016

Binns, R. (2017). 'Algorithmic Accountability and Public Reason', *Philosophy & Technology*

Sharkey, A. (2014) Robots and human dignity: a consideration of the effects of robot care on the dignity of older people, *Ethics and Information Technology* 16, pp. 63–75

Sharkey, A. (2016) Should we welcome robot teachers?, *Ethics and Information Technology* 18, pp. 283–297

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, 352, 1573–1576.