

PROGRAMMI per l'A.A. 2025-2026

1. Facoltà/Istituto/Ciclo di Studio	Filosofia
2. Baccalaureato/Licenza/Dottorato	Licenza
3. Disciplina	Etica del <i>machine learning</i> in IA.
	Ethics of machine learning in AI.
4. Periodo insegnamento (primo semestre, secondo semestre, annuale)	I Semestre
5. N. ore insegnamento	4
6. Settore Scientifico-Disciplinare (SSD), ove già esistente	M-FIL/02-Logica e filosofia della scienza
7. Crediti Formativi CFU/ ECTS	6
8. Cognome e nome del docente	Basti Gianfranco
9. Obiettivi formativi dell'insegnamento/Instructional goals <i>Si dispone di max 3600 caratteri, inclusi gli spazi</i>	Il corso intende offrire un'introduzione approfondita alle questioni etiche sollevate dai sistemi di IA forniti di algoritmi di apprendimento automatico (<i>machine learning</i>) nell'ambito più generale delle questioni etiche legate all'IA
	The course aims to offer an in-depth introduction to the ethical issues raised by AI systems equipped with machine learning algorithms in the more general context of ethical issues related to AI.
10. Prerequisiti/Prerequisites <i>Es. eventuali propedeuticità presenti nel PdS o nelle determinazioni della Facoltà/Istituto; conoscenza di lingue straniere o altro tipo di conoscenze.</i>	Nessuna propedeuticità.
11. Contenuti del corso /Course Contents <i>Si dispone di max 3600 caratteri, inclusi gli spazi</i>	Il corso si divide in Tre Parti. Nella Prima Parte, illustreremo lo sfondo storico e logico-teoretico della problematica, fin dalle origini del programma di ricerca dell'IA, legato al Test di Turing con le sue diramazioni nelle scienze e neuroscienze cognitive, da una parte, e delle reti neurali artificiali, dall'altra. Nella Seconda Parte, illustreremo i principali algoritmi di

	apprendimento automatico supervisionato e non-supervisionato nei sistemi di IA, in particolare gli algoritmi di apprendimento automatico “profondo” (<i>deep-learning</i>) che sono oggi i più efficienti e perciò i più diffusi. Particolare attenzione sarà dedicata ai sistemi di IA generativa, basati sui modelli di apprendimento automatico dei <i>transformers</i>. Nella Terza Parte, illustreremo le diverse questioni etiche sollevate dai sistemi di IA e, in questo ambito, le specifiche problematiche etiche legate agli algoritmi di apprendimento automatico. In particolare, ci soffermeremo sulla questione dell’ineludibile “opacità” del processo decisionale degli algoritmi di apprendimento automatico “profondo” e delle sue ripercussioni etiche e legali, quando questi sistemi sono applicati a decisioni che riguardano la vita e il benessere di singoli e di gruppi. Allo stesso tempo, approfondendo lo sfondo antropologico di tali questioni, discuteremo delle possibili vie di soluzione a queste problematiche, anche in riferimento ai sistemi più avanzati legati alla cosiddetta “IA generativa” tipo Chat-GPT The course is divided into three parts. In Part One, we will illustrate the historical and logical-theoretical background of the problem, since the origins of the AI research program, linked to the Turing Test with its branches in cognitive sciences and neuroscience, on the one hand, and artificial neural networks, on the other. In Part Two, we will illustrate the main supervised and unsupervised machine learning algorithms in AI systems, in particular the deep-learning algorithms that are the most efficient and therefore the most widespread today. Particular attention will be paid to generative AI systems, based on <i>transformers' machine learning models</i>. In Part Three, we will illustrate the different ethical issues raised by AI systems and, in this context, the specific ethical issues related to machine learning algorithms. In particular, we will focus on the question of the inescapable "opacity" of the decision-making process of "deep" machine learning algorithms and its ethical and legal repercussions, when these systems are applied to decisions that affect the life and well-being of individuals and groups. At the same time, deepening the anthropological background of these issues, we will discuss the possible ways of solving these problems, also with reference to the most advanced systems related to the so-called "generative AI" type Chat-GPT
12. Metodo didattico/ Teaching Method <i>Lezioni frontali, laboratori, papers con esposizione elaborati, lavori di gruppo e/o individuali...</i>	Lezioni frontali con supporti multimediali per facilitare l'apprendimento e la discussione sui singoli temi. Lectures with multimedia supports to ease learning and discussion on individual topics.

13. Modalità di verifica dell'apprendimento/Assessment Method <i>Prova orale; scritto (domande aperte, a scelta multipla); prove intermedie.</i>	L'esame sarà scritto con domande sia aperte che a scelta multipla. The exam will be written with both open and multiple-choice questions.
14. Testi di riferimento/Reference Books	- G. BASTI, «Logica aletica, deontica, ontologia formale. Dalla verità ontica all'obbligo deontico». In P. Gherri, G. Basti (Eds.), <i>Logica e Diritto: tra argomentazione e scoperta</i>, Città del Vaticano, Lateran University Press, 2012, pp. 105-270. - M. DAVIS, <i>Il calcolatore universale. Da Leibniz a Turing</i>, 3. Edizione. Milano, Adelphi, 2012. - I. GOODFELLOW, Y. C. BENGIO, <i>Deep Learning</i>, Cambridge UK, Cambridge University Press, 2016. A. DE MAURO, <i>Big data analytics. Analizzare e interpretare dati col machine learning</i>, Online Kindle Edition, 2019. V. C. MÜLLER, «Ethics of Artificial Intelligence and Robotics». In: <i>The Stanford Encyclopedia of Philosophy (Summer 2021 Edition)</i>, Edward N. Zalta (Ed.), URL = <https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/> L. FLORIDI, <i>Etica dell'intelligenza artificiale. Sviluppi, opportunità sfide</i>, Milano, Raffaele Cortina Editore, 2022. G. BASTI, «La sfida etica dell'intelligenza artificiale e il ruolo della filosofia», in <i>Aquinas</i>, LXV, 2 (2022), in stampa. G. BASTI, G. VITIELLO, «Deep learning opacity and the ethical accountability of AI systems. A new perspective». In: R. Giovagnoli, R. Lowe (Eds.), <i>The Logic of Social Practices II. Studies in Applied Philosophy, Epistemology and Rational Ethics (SAPERE)</i>, vol. 43, Berlin-New York, Springer, 2023, in stampa. - G. BASTI. <i>Schemi delle lezioni a uso degli studenti</i>.
15. Criteri per l'assegnazione dell'elaborato finale: tesina, tesi di licenza o dottorale/Thesis assignment criteria	
16. Risultati di apprendimento attesi (secondo i Descrittori di Dublino): Conoscenza e capacità di comprensione <i>Si dispone di max 3600 caratteri, inclusi gli spazi</i>	Conoscenza e capacità di comprensione delle sfide etiche fondamentali dell'IA. Knowledge and understanding of the fundamental ethical challenges of AI.

17. Risultati di apprendimento attesi (secondo i Descrittori di Dublino): <i>Conoscenza e capacità di comprensione applicate</i> <i>Si dispone di max 3600 caratteri, inclusi gli spazi</i>	Capacità di comprendere le sfide etiche fondamentali dei sistemi di IA dotati di apprendimento automatico. Ability to understand the fundamental ethical challenges of AI systems equipped with machine learning.
18. Risultati di apprendimento attesi (secondo i Descrittori di Dublino): <i>Autonomia di giudizio</i> <i>Si dispone di max 3600 caratteri, inclusi gli spazi</i>	Autonomia di giudizio nel saper distinguere fra buona e falsa divulgazione scientifica sulle sfide etiche dell'IA, in base a un discernimento delle fonti. Autonomy of judgment in knowing how to distinguish between good and false scientific dissemination on the ethical challenges of AI, based on a discernment of the sources.
19. Risultati di apprendimento attesi (secondo i Descrittori di Dublino): <i>Abilità comunicative</i> <i>Si dispone di max 3600 caratteri, inclusi gli spazi</i>	Capacità di saper comunicare con chiarezza le principali sfide etiche dell'IA in relazione all'apprendimento automatico. Ability to clearly communicate the main ethical challenges of AI in relation to machine learning.
20. Risultati di apprendimento attesi (secondo i Descrittori di Dublino): <i>Capacità di apprendere</i> <i>Si dispone di max 3600 caratteri, inclusi gli spazi</i>	Capacità di approfondire per proprio conto le principali sfide etiche dell'IA e dell'apprendimento automatico, attingendo direttamente alla letteratura scientifica sull'argomento. Ability to delve into the main ethical challenges of AI and machine learning, drawing directly on the scientific literature on the subject.