

Titolo assegno

La natura della responsabilità nell'AI – Verso un'AI responsabile

Descrizione del progetto

La natura della responsabilità nell'AI – Verso un'AI responsabile

Contesto

L'intelligenza artificiale (AI) offre enormi opportunità per migliorare la vita privata e pubblica, così come il nostro ambiente (si pensi allo sviluppo delle smart cities o ai problemi causati dalle emissioni di carbonio). Sfortunatamente, tali opportunità sono anche associate a significative sfide etiche. La crescente dipendenza dall'AI per modellare le scelte e prendere decisioni, e la graduale riduzione del coinvolgimento umano o addirittura della supervisione su molti processi automatici, pongono questioni urgenti di equità, responsabilità e rispetto dei diritti umani, in particolare della privacy.

Le analisi etiche dell'AI possono aiutare il settore pubblico e privato ad affrontare queste sfide. L'etica dell'AI abbraccia un'ampia gamma di argomenti, dall'attribuzione della responsabilità per gli errori e i fallimenti dell'AI e la spiegazione dei processi dell'AI alla definizione dei principi che guidano l'implementazione dell'AI in una varietà di aree, ad esempio sicurezza, difesa, finanza, sanità e marketing.

Questo progetto, diretto dal Professor Luciano Floridi come “Principal Investigator”, si concentrerà sulla responsabilità dell'AI con gli obiettivi di analizzare i problemi e le soluzioni legali ed etiche esistenti e sviluppare analisi concettuali e normative per affrontarli. Come base specifica del progetto, l'AI responsabile sarà discussa nel contesto dei servizi finanziari.

Argomento

Dato l'uso in rapida crescita dell'AI, è fondamentale garantire che non porti a conseguenze dannose e indesiderate, soprattutto quando coinvolge decisioni o compiti sensibili ed esclude o addirittura preclude la supervisione umana. Questo progetto si concentrerà sulla responsabilità dei servizi di intelligenza artificiale (con un focus sui servizi finanziari). Sebbene vi sia un crescente dibattito e un adeguato livello di comprensione sui danni alla privacy dovuti all'intelligenza artificiale, è stata dedicata un'attenzione limitata all'affrontare questioni relative alla responsabilità per le decisioni basate sull'intelligenza artificiale nel contesto di contesti sempre più basati sui dati e sull'intelligenza artificiale, come quelli dei mercati finanziari. Colmare queste lacune sarà l'obiettivo di questo progetto, che durerà 12 mesi, e consisterà in 2 pacchetti di lavoro.

Pacchetto di lavoro di ricerca 1 (mesi 1- 6) Che cosa intendiamo per responsabilità dell'AI?

Descrizione

Il team di ricerca si concentrerà su questioni teoriche fondamentali riguardanti il concetto di responsabilità nei servizi digitali e su come questo è influenzato dall'intelligenza artificiale. Particolare attenzione sarà data al processo decisionale finanziario basato sui dati e ai quadri esistenti per garantire la

responsabilità in questo contesto. Il team di ricerca si concentrerà sull'attuale lavoro accademico concentrandosi su due ulteriori domande di ricerca: "qual è una comprensione soddisfacente della responsabilità nel caso dei servizi di intelligenza artificiale?" e "possiamo confrontare tale quadro generale sui servizi finanziari basati sull'intelligenza artificiale?"

Prodotti finali

1. Un rapporto su "Accountable AI in Finance, Key Problems and Policy Approaches"
2. Presentazione dei risultati agli eventi Fujitsu, come i forum Fujitsu.
3. Un documento di revisione su "Accountability of AI" da presentare a *Philosophy & Technology* analizzando la letteratura pertinente esistente.
4. Il team potrebbe presentare il documento a ETHICOMP, 2021, la Conferenza internazionale sulle questioni etiche e sociali nelle tecnologie dell'informazione e della comunicazione.

Pacchetto di lavoro di ricerca 2 (mesi 7-12)

Che cosa serve per garantire la responsabilità delle decisioni relative all'AI?

Descrizione

Il team di ricerca identificherà una serie di requisiti necessari per sviluppare un'AI responsabile e a sostegno della privacy, concentrandosi sui servizi finanziari. Il team di ricerca si concentrerà su tre ulteriori domande di ricerca: "i limiti della spiegabilità dell'AI influiscono sulla responsabilità per il suo utilizzo?", "quale livello di trasparenza è necessario per garantire la responsabilità dell'AI?", "come può la protezione della proprietà intellettuale essere bilanciato con la trasparenza necessaria per garantire la responsabilità?". L'analisi si baserà sulla ricerca sia teorica sia empirica, comprendendo casi di studio nei servizi finanziari per confrontare la ricerca e focus group con le parti interessate per valutare i risultati della ricerca. In questa fase, la ricerca adotterà un approccio comparativo, analizzando casi di studio e coinvolgendo le parti interessate in giurisdizioni con diversi quadri di governance legale ed etica, ad esempio Stati Uniti, Regno Unito, UE e Giappone. Questo approccio consentirà di prendere in considerazione diverse strategie politiche e possibili punti di convergenza. Faciliterà inoltre la definizione di requisiti la cui attuazione può essere attuabile anche su scala internazionale, anziché solo nazionale.

Prodotti finali

1. Un rapporto con raccomandazioni valutate su "Requisiti per l'AI responsabile e a supporto della privacy per i servizi finanziari"
2. Presentazione dei risultati agli eventi Fujitsu, come i forum Fujitsu.
3. Workshop di focus group, con le parti interessate del settore per valutare i risultati dello studio con casi d'uso reali.
4. Un documento di ricerca sui requisiti per un'AI responsabile e a sostegno della privacy nei servizi finanziari, da presentare ad *AI & Society* (SpringerNature).

Tempistiche generali e risultati finali del progetto

Ci saranno relazioni trimestrali sui lavori in corso durante i 12 mesi.

- Durante le due fasi (PL1 e PL2)
 - Rapporti interni trimestrali sui lavori in corso
 - Rapporti completi alle conferenze pertinenti e agli eventi pianificati di Fujitsu
 - Focus group con le parti interessate rilevanti
 - Un rapporto su "Accountable AI in Finance, Key Problems and Policy Approaches" che distilla i risultati del documento di revisione

- Un rapporto sui "Requisiti per l'AI a sostegno della privacy e responsabile per i servizi finanziari" che distilla i risultati del documento di ricerca.
- Articolo di revisione da inviare alla rivista internazionale peer-reviewed *Philosophy & Technology* (SpringerNature)
- Documento di ricerca da presentare ad *AI & Society* (SpringerNature)
- Fine del progetto
 - Relazione "executive" che comprende il completamento del progetto.

Description of the Project

The Nature of Accountability in AI – Toward Responsible AI

Outline

Artificial Intelligence (AI) provides huge opportunities to improve private and public life, as well as our environment (consider the development of smart cities or the problems caused by carbon emissions). Unfortunately, such opportunities are also coupled to significant ethical challenges. The growing reliance on AI to shape choices and to make decisions, and the gradual reduction of human involvement or even oversight over many automatic processes, pose pressing issues of fairness, accountability and respect for human rights, particularly privacy.

Ethical analyses of AI can help the private and public sector to meet these challenges. Ethics of AI spans a wide range of topics, from the attribution of responsibility for AI mistakes and failures and the explanation of AI processes to the definition of the principles guiding the deployment of AI in a variety of areas, e.g. security, defence, finance, healthcare, and marketing.

This project, directed by Professor Luciano Floridi as Principal Investigator, will focus on the accountability of AI with the goals of landscaping existing legal and ethical problems and solutions and developing conceptual and normative analyses to address them. As a specific grounding of the project, accountable AI will be discussed in the context of financial services.

Topic

Given the rapidly growing use of AI, it is crucial to ensure that it will not lead to harmful, unintended consequences, especially when it involves sensitive decisions or tasks and excludes or even precludes human supervision. This project will focus on the accountability of AI services (with a focus on financial services). While there is a growing debate and an adequate level of understanding on the privacy harms due to AI, limited attention has been dedicated to addressing questions concerning accountability for AI-based decisions in the context of increasingly datafied and AI-driven contexts, such as financial markets. Filling these gaps will be the goal of this project, which will last 12 months, and consist of 2 work packages.

Research work package 1 (months 1- 6) **What do we mean by accountability of AI?**

Description

The research team will concentrate on fundamental, theoretical questions concerning the concept of accountability in digital services, and how this is impacted by AI. Particular attention will be given to data-driven financial decision-making and existing frameworks to ensure accountability in this context. The research team will concentrate on current scholarly work focusing on two further research questions:

‘what is a satisfactory understanding of accountability in the case of AI services?’ and ‘can we benchmark such general framework on AI-based financial services?’

Deliverables

1. A report on ‘Accountable AI in Finance, Key Problems and Policy Approaches’
2. Presentation of results at Fujitsu events, such as Fujitsu Forums.
3. A review paper on ‘Accountability of AI’ to be submitted to *Philosophy & Technology* analysing the existing relevant literature.
4. The team may present the paper at ETHICOMP, 2021, the International Conference on the Ethical and Social issues in Information and Communication Technologies.

Research work package 2 (months 7-12)

What does it take to ensure the accountability of AI decisions?

Description

The research team will identify a set of necessary requirements to develop accountable, privacy-supporting AI, focusing on financial services. The research team will focus on three further research questions: ‘do the limits of explainability of AI impact the accountability for its deployment?’, ‘what level of transparency is necessary to ensure AI accountability?’, ‘how can the protection of intellectual property be balanced with the transparency necessary to ensure accountability?’. The analysis will rest on both theoretical and empirical research, encompassing case studies in financial services to benchmark the research, and focus groups with relevant stakeholders to evaluate research findings. In this phase, the research will take a comparative approach, analysing case studies and engaging with stakeholders in jurisdictions with different legal and ethical governance frameworks — for example, the US, UK, EU and Japan. This approach will enable the consideration of different policy strategies and possible points of convergence. It will also facilitate the definition of requirements whose implementation may be feasible also on an international, rather than only national, scale.

Deliverables

1. A report with evaluated recommendations on ‘Requirements for Privacy-Supporting, Accountable AI for Financial Services’
2. Presentation of results at Fujitsu events, such as Fujitsu Forums.
3. Focus group workshops, with industry stakeholders to evaluate the findings of study with real use cases.
4. A research paper on the requirements for privacy-supporting, accountable AI in financial services, to be submitted to *AI & Society* (SpringerNature).

Project overall timeline and deliverables

There will be quarterly reports of work in progress during the 12 months.

- During the two phases (WP1 and WP2)
 - Quarterly internal reports of work in progress
 - Full reports at relevant conferences and planned Fujitsu’s events
 - Focus groups with relevant stakeholders
 - A report on ‘Accountable AI in Finance, Key Problems and Policy Approaches’ distilling the findings of the review paper
 - A report on ‘Requirements for Privacy-Supporting, Accountable AI for Financial Services’ distilling the findings of the research paper.



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
DIPARTIMENTO DI SCIENZE GIURIDICHE

- Review article to be submitted to the peer-reviewed, international journal *Philosophy & Technology* (SpringerNature)
 - Research paper to be submitted to *AI & Society* (SpringerNature)
- End of project
 - Executive report encompassing project completion.