Application Package

Poste de professeure ou professeur en gestion de l'intelligence artificielle

Offre 07823

Carlos Denner dos Santos

AI Scientist
Montreal, QC, Canada

carlosdenner@gmail.com
+1 (438) 836-4116

Département des systèmes d'information et méthodes quantitatives de gestion (SIMQG)
École de gestion
Université de Sherbrooke

17 novembre 2025

Table des matières

CARLOS DENNER DOS SANTOS

AI Scientist ◇ Montreal, QC, Canada
carlosdenner@gmail.com ◇ +1 (438) 836-4116
LinkedIn ◇ ResearchGate ◇ Google Scholar

RESEARCH INTERESTS

LLM security and prompt injection defense, affective AI and human–LLM interaction, sentiment analysis and algorithmic feelings, AI governance and ethical AI systems, retrieval-augmented generation (RAG) and hallucination mitigation, digital innovation theory and reversible control, AI-driven digital inclusion, open-source software ecosystems, IT governance and digital transformation.

EDUCATION

University of Nottingham, England — *December 2010 - December 2011*
PostDoc in Computer Science
Supervisor: George Kuk

University of São Paulo, Brazil — *August 2009 - July 2011*
PostDoc in Computer Science
Supervisor: Fabio Kon

Southern Illinois University, United States — *August 2005 - July 2009*
PhD in Information Systems
Supervisor: John Pearson

Federal University of Minas Gerais, Brazil — *February 2003 - February 2005*
MSc in Management

State University of Minas Gerais, Montes Claros — *February 1999 - December 2002*
BSc in Business

Technical School of Montes Claros, Brazil — *February 1997 - December 1999*
Technical Degree in Data Processing

PROFESSIONAL EXPERIENCE (1999–2025)

AI Expert — *July 2025 - Present*
Videns AI — Montreal, Canada

Data Scientist Consultant — *November 2021 - January 2025*
Bell Canada — Montreal, Canada

Research Associate — *May 2021 - April 2023*
École de Technologie Supérieure — Montreal, Canada

Co-Investigator — *February 2020 - Present*
Jooay App, McGill University — Montreal, Canada

Assistant Researcher — *November 2019 - February 2020*
CHU Sainte-Justine Research Center — Montreal, Canada

Research Associate — *June 2019 - May 2021*
Université du Québec à Montréal (UQAM) — Montreal, Canada

Board Member — *June 2017 - December 2018*
FINATEC (Foundation for Technological Projects) — Brasília, Brazil

Consultant — *November 2016 - February 2017*
Tribunal de Contas da União (Federal Court of Accounts) — Brasília, Brazil

Adjunct Director — *June 2016 - December 2020*
PhD Program in Management, University of Brasília — Brasília, Brazil

Associate Director — *January 2016 - December 2017*
Center for Technology Development (CDT), University of Brasília — Brasília, Brazil

Consultant (Open Source Software Governance) — *February 2014 - June 2016*
Ministry of Planning, Government of Brazil — Brazil

Consultant — *November 2013 - February 2014*
Ministry of Planning, Government of Brazil — Brazil

Founder & Manager — *January 2013 - December 2023*
Socie-Dados Research Lab, University of Brasília — Brasília, Brazil

Associate Professor — *December 2011 - December 2023*
University of Brasília — Brasília, Brazil

Postdoctoral Researcher — *December 2010 - December 2011*
University of Nottingham — Nottingham, UK

Collaborating Professor — *August 2010 - January 2011*
University of São Paulo — São Paulo, Brazil

Postdoctoral Researcher — *August 2009 - August 2011*
University of São Paulo — São Paulo, Brazil

Project Advisor (SoftwarePublico.gov.br) — *July 2009 - December 2010*
CTI Renato Archer — Campinas, Brazil

Graduate Instructor — *January 2008 - May 2008*
Southern Illinois University Carbondale — Carbondale, IL, USA

Instructor (Professional Training) — *June 2005*
FAPEMIG, Minas Gerais Research Foundation — Brazil

Part-Time Lecturer — *August 2004 - July 2005*
Universidade FUMEC — Belo Horizonte, Brazil

Consultant (IT) — *May 2004 - December 2004*
Bioquímica e Química do Brasil Ltda. — Belo Horizonte, Brazil

Substitute Professor — *March 2004 - August 2005*
Federal University of Minas Gerais — Belo Horizonte, Brazil

Computer Programmer — *August 2002 - February 2003*
Universidade Estadual de Montes Claros — Montes Claros, Brazil

Professor — *August 2001 - February 2003*
Fundação Educacional Montes Claros — Montes Claros, Brazil

Network Administrator / Computer Programmer — *March 1999 - January 2003*
Unimed Montes Claros — Montes Claros, Brazil

ACADEMIC SUPERVISIONS

Postdoctoral Supervisions

Pedro Jacome de Moura Junior *2023*
Universidade de Brasília

Almir Oliveira Junior *2018*
Instituto de Pesquisa Econômica Aplicada (IPEA), Universidade de Brasília

Fabio Buiati *2017*
Centro de Apoio ao Desenvolvimento Tecnológico, Universidade de Brasília

PhD Supervisions

Anna Carolina Ribeiro *2022*
Risk management in public administration · Universidade de Brasília

Claudia Tolentino Santos *2021*
Pension governance and financial performance · Universidade de Brasília

Pablo Péron *2021*
Organizational survival post-incubation (AI-driven analysis) · Universidade de Brasília

Gustavo Alves *2021*
IT governance in the public sector · Universidade de Brasília

Patrícia Almeida *2019–2024*
AI governance and digital government · Universidade de Brasília

Deise Goulart *2021–Present*
NGO and public sector efficiency · Universidade de Brasília

Isabela Ferraz *2019*
Virtual community governance · Universidade de Brasília

Silvia Satiko Onoyama *2018*
Mechanisms for effective IT governance · Universidade de Brasília

Paulo Meirelles (co-supervisor) *2013*
Open-source software metrics monitoring · Universidade de São Paulo

MSc Supervisions

Leonel Cerqueira Santos *2016*
Non-operational mechanisms in IT governance effectiveness · Universidade de Brasília

Daniel Shim de Sousa Esashika *2016*
Influence of sponsors on open-source community structures · Universidade de Brasília

Isadora Vergara *2015*
Adaptive management vs. environmental determinism in blogs · Universidade de Brasília

Leonardo Oliveira *2015*
Consequences of innovation adoption · Universidade de Brasília

Juliana Miranda *2015*
Environmental and organizational variables in startup performance · Universidade de Brasília

Luiz Fernando Silva Pinto *2015*
Factors influencing effort and contributions on Wikipedia · Universidade de Brasília

Selected Undergraduate Research Supervisions

2020–2021: Thiago Lopes, Sara Andrade, Rayssa Lorrane, Rafael Azevedo Lima, Eduardo Martins, Edilson Niehues Rodrigues Lima
2018–2019: Valesca Scarlat C. da Fonseca, Raissa Paiva Pires, Vitor Gabriel G. da Silva, Saulo Barros de Melo, Ateldy Brasil Filho
2012–2014: Leonardo Alves dos Santos, André S. R. de Souza Marques, Wilton da Silva Rodrigues, Thiago F. Figueiredo, Lucas B. Cusinato Rodrigues
2009–2010: Rafael S. Suguiura (USP), Marcos Bonci (USP)

Selected Specialization/MBA & Undergraduate Final Projects

2021: Luiz Felipe Pimenta de Araujo (Serious games in teaching business strategy)
2014–2016: 15+ supervisions in business strategy and management topics

PUBLICATIONS

Journal Articles

[1] P. G. R. d. Almeida and C. D. d. Santos. **Artificial intelligence governance: understanding how public organizations implement IT**. *Government Information Quarterly* **42**.1 (2025), pp. 102003–102003.

[2] E. Mahmoudi, M. Movahed, A. Majnemer, C. D. d. Santos, G. Backlin, A. Siou, S. Glegg, L. Pritchard, J. McCabe, and K. Shikako-Thomas. **Gamification strategies that promote leisure participation in children and youth with disabilities**. *Disability and Rehabilitation Assistive Technology* (2025), pp. 1–21.

[3] S. Mori, C. Santos, and I. Ferraz. **Efeitos dos mecanismos de governança na capacidade absortiva de projetos de colaboração internacional: um estudo de caso na organização Embrapa**. *Revista Brasileira de Inovação 24, e* (2025).

[4] S. Mori, C. Santos, and I. Ferraz. **Effects of governance mechanisms on the absorptive capacity of international collaboration projects: a case study in the company Embrapa**. *Revista Brasileira de Inovação 24, e* (2025).

[5] L. F. P. d. Araújo, C. D. d. Santos, I. N. Ferraz, and P. P. d. Paula. **Ensinando estratégia empresarial usando jogos sérios na graduação: estado atual e potencialidades futuras**. *Gestão org* **21**.1 (2024). Open Access.

[6] E. Mahmoudi, P. Y. Yoo, A. Chandra, R. Cardoso, C. D. d. Santos, A. Majnemer, and K. Shikako-Thomas. **Gamification in mobile apps for children with disabilities: scoping review**. *JMIR Serious Games* **12** (2024). Open Access, e49029–e49029.

[7] P. J. d. Moura, C. D. d. Santos, C. G. P. Bellini, and J. J. L. Dias. **The Over-Concentration of innovation and Firm-Specific knowledge in the artificial intelligence industry**. *Journal of the Knowledge Economy* **15**.4 (2024), pp. 20547–20577.

[8] P. P. d. Paula and C. D. d. Santos. **Vulnerabilidade INICIAL PÓS-INCUBAÇÃO: PREVENDO a SOBREVIVÊNCIA ORGANIZACIONAL COM APRENDIZAGEM DE MÁQUINA**. *Revista Gestão e Desenvolvimento* **21**.1 (2024). Open Access, pp. 28–50.

[9] D. A. d. Silva, C. D. d. Santos, and I. N. Ferraz. **Como promover a colaboração em plataformas de dados abertos?** *Revista de Ciências da Administração* **26**.66 (2024). Open Access, pp. 1–30.

[10] E. Mahmoudi, P. Y. Yoo, A. Chandra, R. Cardoso, C. D. d. Santos, A. Majnemer, and K. Shikako-Thomas. **Gamification in mobile apps for children with disabilities: scoping review (Preprint)** (2023). Open Access.

[11] P. P. d. Paula, C. D. d. Santos, and F. F. Couto. **Organizational survival of Technology-Based companies after incubation: a qualitative comparative explanation**. *Review of Business Management* **25**.4 (2023). Open Access, pp. 498–515.

[12] G. d. F. Alves and C. D. d. Santos. **The diffusion of innovations under normative induction in Brazil**. *RAUSP Management Journal* **57**.2 (2022). Open Access, pp. 149–164.

[13] A. Etkkali, P.-N. Placide, and C. D. d. Santos. **Ict governance performance implications in higher education: a systematic review of the literature** (2022), pp. 102–107.

[14] I. N. Ferraz and C. D. d. Santos. **Transformation OF FREE AND OPEN SOURCE SOFTWARE DEVELOPMENT PROJECTS: GOVERNANCE BETWEEN THE CATHEDRAL AND BAZAAR**. *Revista de Administração de Empresas* **62**.1 (2022). Open Access.

[15] L. F. S. Pinto, C. D. d. Santos, and S. S. Onoyama. **Individual factors that influence effort and contributions on Wikipedia**. *arXiv (Cornell University)* (2022). Open Access.

[16] C. D. d. Santos and J. Alves. **Home advantage in the Brazilian elite football: verifying managers' capacity to outperform their disadvantage**. *arXiv (Cornell University)* (2022). Open Access.

[17] C. D. d. Santos, I. Castro, G. Kuk, S. S. Onoyama, and M. F. Moreira. **Bucking the trend: an agentive perspective of managerial influence on blogs attractiveness**. *arXiv (Cornell University)* (2022). Open Access.

[18] C. D. d. Santos, M. Moreira, A. Panis, and D. Alves. **Crowdfunding for entrepreneurial education**. *International Journal of Information and Communication Technology Education* **18**.1 (2022). Open Access, pp. 1–13.

[19] P. G. R. d. Almeida, C. D. d. Santos, and J. S. Farias. **Artificial intelligence regulation: a framework for governance**. *Ethics and Information Technology* **23**.3 (2021), pp. 505–525.

[20] I. N. Ferraz and C. D. d. Santos. **Organization of free and open source software projects: in-between the community and traditional governance**. *Brazilian Business Review* **18**.3 (2021). Open Access, pp. 334–352.

[21] D. A. d. Silva, J. A. D. SILVA, G. d. F. Alves, and C. D. d. Santos. **Gestão de riscos no setor público: revisão bibliométrica e proposta de agenda de pesquisa**. *Revista do Serviço Público* **72**.4 (2021). Open Access, pp. 824–854.

[22] P. Y. Yoo, M. Movahed, I. Rue, C. D. d. Santos, A. Majnemer, and K. Shikako-Thomas. **Changes in use of a leisure activity mobile app for children with disabilities during the COVID-19 pandemic: retrospective study**. *JMIR Pediatrics and Parenting* **5**.1 (2021). Open Access, e32274–e32274.

[23] D. Esashika and C. D. d. Santos. **The influence of sponsors on organizational structure of free software communities**. *arXiv (Cornell University)* (2020). Open Access.

[24] I. Ferraz and C. Santos. **Transformação de projetos de desenvolvimento de software livre: Uma governança entre a catedral e o bazar**. *Revista de Administração de Empresas 62, e* (2020).

[25] A. C. M. L. Ribeiro and C. D. d. Santos. **Isso não é uma pirâmide: revisando o modelo clássico de dado, informação, conhecimento e sabedoria**. *Ciência da Informação* **49**.2 (2020). Open Access.

[26] A. C. M. L. Ribeiro and C. D. d. Santos. **This is not a pyramid: revising the data, information, knowledge and wisdom classical model** (2020). Open Access.

[27] C. R. L. Kohler, C. D. d. Santos, and M. Bursztyn. **Understanding environmental terrorism in times of climate change: implications for asylum seekers in germany**. *Research in Globalization* **1** (2019). Open Access, pp. 100006–100006.

[28] L. F. d. Oliveira and C. D. d. Santos. **Intended AND UNINTENDED CONSEQUENCES OF INNOVATION ADOPTION: OPEN GOVERNMENT DATA ADOPTION BY THE FEDERAL DISTRICT OF BRAZIL**. *REAd Revista Eletrônica de Administração (Porto Alegre)* **25**.1 (2019). Open Access, pp. 1–25.

[29] L. Oliveira and C. Santos. **CONSECUENCIAS PREVISTAS Y NO PREVISTAS DE LA ADOPCIÓN DE INNOVACIÓN: ADOPCIÓN DE DATOS ABIERTOS POR PARTE DEL GOBIERNO DE DISTRITO FEDERAL DE BRASIL**. *REAd. Revista Eletrônica de Administração (Porto Alegre) 25, 1-25* (2019).

[30] L. Oliveira and C. Santos. **CONSEQUÊNCIAS PRETENDIDAS E NÃO PRETENDIDAS DA ADOÇÃO DE INOVAÇÕES: A ADOÇÃO DE DADOS ABERTOS PELO GOVERNO DO DISTRITO FEDERAL DO BRASIL**. *REAd. Revista Eletrônica de Administração (Porto Alegre) 25, 1-25* (2019).

[31] T. L. Gelenske, J. S. Farias, and C. D. d. Santos. **The PERCEIVED RISK AND TRUST IN THE BANK'S BRAND ON THE MOBILE BANKING USER VISION**. *Reuna* **23**.2 (2018). Open Access, pp. 1–22.

[32] L. F. d. Oliveira and C. D. d. Santos. **Public value innovation** (2018), pp. 1–9.

[33] P. P. d. Paula, J. S. Farias, and C. D. d. Santos. **Uma ANÁLISE SOBRE STARTUPS DE BASE TECNOLÓGICA a PARTIR DA LITERATURA DO PERÍODO 2008-2017**. *Anais do ... EGEPE* (2018). Open Access.

[34] L. F. S. Pinto and C. D. d. Santos. **Motivations of crowdsourcing contributors**. *Innovation & Management Review* **15**.1 (2018). Open Access, pp. 58–72.

[35] D. Esashika and C. D. d. Santos. **The influence of sponsors on organizational structure of free software communities** (2017), pp. 265–272.

[36] L. F. S. Pinto, S. S. O. Mori, and C. D. d. Santos. **Inovação na contramão: uma bibliometria dos artigos relativos aos usuários que inventam**. *Revista de Administração* **15**.3 (2017). Open Access, pp. 74–94.

[37] C. D. d. Santos. **Changes in free and open source software licenses: managerial interventions and variations on project attractiveness**. *Journal of Internet Services and Applications* **8**.1 (2017). Open Access.

[38] L. Santos and C. S. Jr. **Um estudo sobre o impacto dos mecanismos não operacionais na efetividade da governança detecnologia da informação pública**. *Revista de Administração (São Paulo) 52 (3), 256-267* (2017).

[39] L. C. Santos and C. D. d. Santos. **A study on the impact of non-operational mechanisms on the effectiveness of public information technology governance**. *Revista de Administração* **52**.3 (2017). Open Access, pp. 256–267.

[40] J. Q. Miranda, C. D. d. Santos, and A. T. Dias. **A INFLUÊNCIA DAS VARIÁVEIS AMBIENTAIS e ORGANIZACIONAIS NO DESEMPENHO DE STARTUPS**. *REGEPE Entrepreneurship and Small Business Journal* **5**.1 (2016). Open Access, pp. 28–65.

[41] I. V. Castro and C. D. d. Santos. **"o que gerencio e de quem dependo?": determinantes da ação de blogueiros**. *Revista de Administração Contemporânea* **19**.4 (2015). Open Access, pp. 486–507.

[42] L. Carletti, T. Coughlan, J. Christensen, E. M. Gerber, G. Giannachi, S. Schutt, R. Sinker, and C. D. d. Santos. **Structures for knowledge co-creation between organisations and the public** (2014), pp. 309–312.

[43] J. S. Farias, S. Sanders, C. D. d. Santos, and K. Rozzett. **Aceitação de tecnologia em terminais de autosserviço aeroportuários: explorando os efeitos dos moderadores idade, experiência e gênero** (2014). Open Access, pp. 66–77.

[44] G. Heindrickson and C. S. Jr. **Governança de ti em organizações públicas: como a efetividade percebida se relaciona com três mecanismos clássicos**. *JISTEM-Journal of Information Systems and Technology Management 11 (2), 297-326* (2014).

[45] G. Heindrickson and C. D. d. Santos. **Information technology governance in public organizations: how perceived effectiveness relates to three classical mechanisms**. *Journal of Information Systems and Technology Management* **11**.2 (2014). Open Access, pp. 297–326.

[46] C. d. S. Jr. **Edição temática: Software Livre/EDITOR's SPACE: Special issue: Free Software**. *Revista Electronica de Sistemas de Informaçao 13 (2), 1* (2014).

[47] C. d. S. Jr. **EDITOR's SPACE Special issue: Free Software**. *Revista Eletrônica de Sistemas de Informação 13 (2)* (2014).

[48] C. D. d. Santos. **Editorial**. *Revista Eletrônica de Sistemas de Informação* **13**.2 (2014). Open Access.

[49] N. Williams. **Global entrepreneurship: case studies of entrepreneurial firms operating around the world**. *International Journal of Entrepreneurial Behaviour & Research* **21**.4 (2014), pp. 642–643.

[50] I. T. V. M. N. Castro and C. D. d. Santos. **Popularidade em blogs: composição, causas e consequências**. *Revista Brasileira de Administração Científica* **4**.2 (2013). Open Access, pp. 185–198.

[51] L. Franco and C. Denner. **Adoção de Práticas de Gestão de Segurança da Informação na Administração Pública Federal (Adoption of Information Security Management Practices in the Federal Public ...** *SSRN Electronic Journal* (2013).

[52] L. Franco and C. Denner. **Adoção de Práticas de Gestão de Segurança da Informação na Administração Pública Federal (Adoption of Information Security Management Practices in the Federal Public Administration)**. *SSRN Electronic Journal* (2013).

[53] C. D. d. Santos, G. Kuk, F. Kon, and J. Pearson. **The attraction of contributors in free and open source software projects**. *The Journal of Strategic Information Systems* **22**.1 (2012). Open Access, pp. 26–45.

[54] V. A. d. Santos, A. Goldman, and C. D. d. Santos. **Uncovering steady advances for an extreme programming course**. *CLEI electronic journal* **15**.1 (2012). Open Access.

[55] P. Meirelles, C. D. d. Santos, J. M. d. Miranda, F. Kon, A. Terceiro, and C. Chávez. **A study of the relationships between source code metrics and attractiveness in free software projects** (2010), pp. 11–20.

[56] C. D. d. Santos and K. M. Nelson. **Motivation TO CREATE FREE AND OPEN SOURCE PROJECTS AND HOW DECISIONS IMPACT SUCCESS**. *Revista Eletrônica de Sistemas de Informação* **9**.2 (2010), pp. 4–4.

[57] C. D. d. Santos. **Understanding partnerships between corporations and the open source community: a research gap**. *IEEE Software* **25**.6 (2008), pp. 96–97.

[58] S. Brookhart, C. d. Flora, and C. D. d. Santos. **Bookshelf**. *IEEE Software* **24**.1 (2007), pp. 92–94.

[59] C. D. d. Santos and M. A. Gonçalves. **Análise da substituição de um software proprietário por umSoftware livre sob a ótica do custo total de Propriedade:Estudo de caso do setor de peças automobilísticas**. *DOAJ (DOAJ: Directory of Open Access Journals)* (2006). Open Access.

Conference Papers

[1] "Estudos de Aston e a Teoria Formal da Diferenciação Estrutural Revisitados: Os Resultados Ainda se Aplicam?" 2025.

[2] A. C. M. L. Ribeiro, G. Demo, and C. D. d. Santos. "Grupo FOCAL: APLICAÇÕES NA PESQUISA NACIONAL EM ADMINISTRAÇÃO". *Dialnet (Universidad de la Rioja)*. 2021.

[3] C. D. d. Santos, M. A. R. Dantas, R. A. d. Carvalho, H. L. C. Júnior, and L. A. Amaral. "Tecnologias e dados abertos para inovação em governo". 2021.

[4] C. D. d. Santos, I. d. Castro, S. S. Onoyama, and M. F. Moreira. "Bucking the trend: an agentive perspective of managerial influence on Blog'S attractiveness". *Proceedings of the ... Annual Hawaii International Conference on System Sciences/Proceedings of the Annual Hawaii International Conference on System Sciences*. Open Access. 2020.

[5] A. Panis, C. D. d. Santos, and J. B. S. Silva. "Crowdfunding in higher education: a classroom exercise for knowledge application and transfer". *AIS Electronic Library (AISeL) (Association for Information Systems)*. 2019.

[6] D. A. Silva and C. D. d. Santos. "Open government data: from transparency to social participation". *AIS Electronic Library (AISeL) (Association for Information Systems)*. 2019.

[7] G. d. F. Alves and C. D. d. Santos. "The day after adoption: managerial innovation abandonment and recycling". *AIS Electronic Library (AISeL) (Association for Information Systems)*. Open Access. 2018.

[8] C. d. Santos. "Whose Stake in Open Source Software Projects." *AMCIS*. 2018.

[9] B. Henrique and C. d. S. Júnior. "As consequências organizacionais das escolhas estratégicas das coalizões dominantes". *Diálogos Interdisciplinares 6 (1), 86-101*. 2017.

[10] L. F. d. Oliveira and C. D. d. Santos. "Inovações no setor público : uma abordagem teórica sobre os impactos de sua adoção". 2017.

[11] L. OLIVIERA. "SANTOS JÚNIOR, Carlos Denner". *Inovações no setor público: uma abordagem teórica sobre os impactos de sua ...* 2017.

[12] L. Santos and C. D. Santos. "Un estudio acerca de la influencia de los mecanismos no operacionales en la efectividad de lagobernanza de tecnología de la información pública". *Sao Paulo: Scielo. Obtenido de http://www. scielo. br/scielo. php*. 2017.

[13] L. F. d. Oliveira and C. D. d. Santos. "The two sides of the innovation coin". *Americas Conference on Information Systems*. 2016.

[14] T. Gelenske, J. Farias, and C. D. S. Junior. "A Relação entre o risco percebido e a confiança na marca do banco na ótica de usuários de mobile banking". *Seminários de Administração 27, 1-17*. 2015.

[15] D. Hastenreiter and C. D. d. Santos. "Impactos da escolha da licença na dinâmica de desenvolvimento de software livre". *Americas Conference on Information Systems*. 2015.

[16] L. F. S. Pinto and C. D. d. Santos. "Motivações dos contribuidores de crowdsourcing full paper". 2015.

[17] C. D. d. Santos and L. F. S. Pinto. "Motivações dos contribuidores de crowdsourcing". *Americas Conference on Information Systems*. 2015.

[18] I. Castro, J. Farias, and C. Júnior. "Redes Sociais Virtuais-Uma investigação de abordagens metodológicas de pesquisa". *CIAIQ*. 2014.

[19] N. Santos, C. d. Sousa, C. d. Santos, and R. Santos. "Business Incubation and the Pipeway Business Case in Brazil". *Global Entrepreneurship, 1-10*. 2014.

[20] C. Chavez, A. Terceiro, P. Meirelles, C. S. Jr, and F. Kon. "Free/libre/open source software development in software engineering education: Opportunities and experiences". *Fórum de Educaçao em Engenharia de Software (CBSoft'11-SBES-FEES)*. 2011.

[21] C. S. Jr, G. Kuk, F. Kon, and R. Suguiura. "The Inextricable Role of Organizational Sponsorship for Open Source Sustainability". *Proceedings of SOS*. 2011.

[22] P. Meirelles, F. Kon, and C. S. Jr. "Semi-Automatic Evaluation of Free Software Projects: A Source Code Perspective". *Salvador, Brazil, 42*. 2011.

[23] C. D. d. Santos, M. B. Cavalca, F. Kon, J. M. Singer, V. Ritter, D. Regina, and T. Tsujimoto. "Intellectual property policy and attractiveness". *Conference on Computer Supported Cooperative Work*. 2011.

[24] C. Melo, C. Jr, G. Ferreira, and F. Kon. "Um estudo exploratório dos fatores associados ao estímulo do aprendizado em times ágeis na indústria". *Experimental Software Engineering Latin American Workshop 7, 80-98*. 2010.

[25] C. D. d. Santos. "Atratividade de projetos de software livre: importância teórica e estratégias para administração". *Sociedad (University of Buenos Aires)*. Open Access. 2010.

[26] C. D. d. Santos, F. Kon, and J. Pearson. "Attractiveness of free and open source software projects". *AIS Electronic Library (AISeL) (Association for Information Systems)*. 2010, p. 105.

[27] C. S. Jr. "Open source software projects' attractiveness, activeness, and efficiency as a path to software quality: an empirical evaluation of their relationships and causes". *Southern Illinois University at Carbondale*. 2009.

[28] C. S. Jr and M. Gonçalves. "A resource-based explanation for the Apache consistent dominance in the web-server industry". *Associação Nacional dos Programas de Pós-Graduação em Administração (ANPAD)*. 2008.

[29] K. Nelson and C. Santos. "Attractiveness of Open Source Projects: A Path to Software Quality". *AMCIS*. 2007.

[30] C. D. d. Santos and M. A. Gonçalves. "Análise da substituição de um software proprietário por um software livre sob a ótica do TCO (Custo total de Propriedade):Estudo de caso do setor de peças automobilísticas". *Americanae (AECID Library)*. Open Access. 2005.

[31] J. AMÂNCIO, C. S. JÚNIOR, and M. GONÇALVES. "Avaliação de sistemas de informações". *CONGRESSO ANUAL DE TECNOLOGIA DA INFORMAÇÃO. Anais CATI, 1-14*. 2004.

Book Chapters

[1] A. C. M. L. Ribeiro, P. C. Ferreira, and C. D. d. Santos. **Automação e mercado de trabalho: análise da literatura e evidências empíricas**. *IPEA eBooks*. Open Access. 2024, pp. 217–253.

[2] S. Barbalho, G. Ghesti, S. Carvalho, C. d. Santos, and . AR Martin. **Capítulo 2 A Gestão da Inovação na Universidade de Brasília**. *BOAS PRÁTICAS DE GESTÃO EM NÚCLEOS DE INOVAÇÃO TECNOLÓGICA: Experiências ...* 2019.

[3] L. F. d. Oliveira and C. D. d. Santos. **Open innovation in the public sector**. *Springer proceedings in complexity*. 2018, pp. 458–466.

[4] C. D. d. Santos, M. A. Gonçalves, and F. Kon. **Apache sustained competitive advantage in the web server industry**. *IGI Global eBooks*. Open Access. 2011, pp. 259–273.

Datasets

[1] P. P. d. Paula, C. D. d. Santos, and F. F. Couto. **Supplementary data - organizational survival of Technology-Based enterprises after incubation: a qualitative comparative explanation**. Harvard Dataverse. Research Dataset. 2023.

[2] I. V. Castro and C. D. d. Santos. **"what do i manage and on whom do i depend": determinants of bloggers' actions**. Research Dataset. 2022.

[3] L. C. Santos and C. D. d. Santos. **A study on the impact of non-operational mechanisms on the effectiveness of public information technology governance**. Research Dataset. 2021.

Open Peer Reviews

[1] C. D. d. Santos. **Review of: "Artificial intelligence and organizational change"** (2023). Open Access.

Under Review & In Preparation

Denner, C. D. (et al.). Building an LLM Firewall: A Multi-Phase Defense Against Prompt Injection – From Patent Landscape to Deployable Input-Side Guardrails. *Communications of the ACM* – Under review.

Denner, C. D. (et al.). Reversible Control as a Digital Innovation Theory: Changing the Control–Learning Relationship through Knowledge Deployment Capability. *MISQ Theory & Review track* – Manuscript in preparation.

Denner, C. D. LLMs, Sentiment Analysis, and Algorithmic Feelings: Toward a Theory of Affective Loops in Human–LLM Interaction. *Academy of Management Review* – Manuscript in preparation.

Denner, C. D. (et al.). The AI Recommendation System of Jooay.com: Enhancing Digital Inclusion for Children and Youth with Disabilities. Manuscript in preparation.

Denner, C. D. (et al.). Evaluating and Mitigating Hallucinations in Retrieval-Augmented Generation (RAG): An Experimental Framework. Manuscript in preparation.

RESEARCH GRANTS

FAP-DF Demanda Espontânea *2016–2021*
Principal Investigator
Research Support Foundation of Distrito Federal

CAPES Pesquisador Júnior *2018*
Principal Investigator
Young Investigator Grant, Coord. for Improvement of Higher Education Personnel

CNPq Edital Universal *2013–2015*
Principal Investigator
Universal Research Grant, National Council for Scientific and Technological Development

AWARDS & HONORS

Horizon Institute Postdoctoral Fellowship *2011–2012*
Horizon Digital Economy Research Institute, UK · University of Nottingham

FAPESP Postdoctoral Fellowship *2009–2010*
São Paulo Research Foundation, Brazil · University of São Paulo

Pontikes Research Center Fellowship *2007*
Southern Illinois University · Open-source community partnership study

John M. Fohr Memorial Scholarship *2007*
Southern Illinois University Foundation · Excellence in Management studies

Fulbright/CAPES PhD Fellowship *2005–2009*
Fulbright Commission & CAPES, Brazil · Full fellowship for PhD studies in the USA

CNPq MSc Fellowship *2003–2005*
CNPq, Brazil · Graduate scholarship for M.Sc. studies in Management

Honorable Mentions *2013, 2015*
Two research papers received honorable mention at EnANPAD (Brazilian national conference)

PROFESSIONAL SERVICE

Editorial Board Member *2015–2025*
Journal of Global Information Technology Management

Ad-hoc Reviewer (Funding Agencies) *2015–Present*
Research grant proposals for CAPES (2015) and FAP-DF (2015–Present)

Board Member *2017–2018*
Foundation for Technological Projects (FINATEC), Brasília · University-industry partnerships

Consultant (Data Analytics) *2016–2017*
Tribunal de Contas da União (Federal Audit Court), Brazil · Education data analysis

Consultant (Open Source Governance) *2014–2016*
Ministério do Planejamento – softwarepublico.gov.br · Platform governance and licensing

Consultant (IT Management) *2013–2014*
Ministério do Planejamento, Brazil · IT workforce estimation and decision support

PhD Defense Committee Member *2013–2023*
Multiple dissertation committees at Universidade de Brasília and other institutions

Academic Hiring Committee Member *Various years*
University faculty hiring panels and graduate scholarship selection committees

Peer Reviewer (14 reviews, 2021–2025)
· Future Business Journal (2025)
· Heliyon – Cell Press (2024)
· Public Management Review (2023, 2 reviews)
· Journal of Global Information Technology Management (2022, 2023, 3 reviews)
· Journal of Open Innovation: Technology, Market, and Complexity (2023)
· Journal of Forensic Psychology Research and Practice (2023)
· IEEE Access (2022, 2 reviews)
· ACM Transactions on Software Engineering and Methodology (TOSEM) (2022)
· Revista de Administração Contemporânea (2022)

Conference Reviewer
AMCIS, ECIS, ICIS, AoM, ANPAD (EnANPAD), and other information systems and management conferences

PROFESSIONAL MEMBERSHIPS

Association for Information Systems (AIS) *2015 - Present*

Brazilian Academy of Management (ANPAD) *2010 - Present*

Academy of Management (AOM) *2016 - Present*

Lettre de motivation

Sherbrooke, le [date]

Objet : Candidature au poste de professeure ou professeur en gestion de l'intelligence artificielle (offre 07823)

Madame, Monsieur,

Je vous soumets ma candidature au poste de professeure ou professeur en gestion de l'intelligence artificielle (offre 07823) au Département des systèmes d'information et méthodes quantitatives de gestion (SIMQG) de l'École de gestion de l'Université de Sherbrooke.

Titulaire d'un doctorat en systèmes d'information au sein d'une faculté de gestion, complété par des stages postdoctoraux en informatique et en statistique, je travaille depuis plus de vingt ans à l'intersection de l'analytique, des systèmes d'information et de la stratégie. Mes projets récents portent sur la gouvernance et la régulation de l'IA, la gestion de portefeuilles de projets d'IA, ainsi que sur l'ingénierie et l'exploitation de systèmes d'IA à grande échelle, incluant des modèles de type LLM et des systèmes de recommandation.

Sur le plan scientifique, j'ai publié notamment dans Ethics and Information Technology (« Artificial Intelligence Regulation : a framework for governance ») et dans Government Information Quarterly (« Artificial intelligence governance : Understanding how public organizations implement it »). Ces travaux proposent, d'une part, un cadre intégrateur pour la régulation de l'IA, et d'autre part, une étude empirique de la mise en œuvre de la gouvernance de l'IA dans 28 organisations publiques sur cinq continents. Ils s'inscrivent directement dans les thématiques du poste, en particulier la gouvernance et la gestion des risques de l'IA ainsi que la sécurité et la résilience des systèmes d'IA.

Parallèlement, j'ai conçu et déployé des systèmes analytiques et d'IA dans des secteurs variés (télécommunications, énergie, santé, économie sociale), ce qui me donne une vision très concrète des enjeux d'AIOps/MLOps et de cycle de vie de l'IA en contexte organisationnel. Mes recherches actuelles portent sur la sécurité des LLM (défense contre la prompt injection, détection des hallucinations dans les systèmes RAG), l'intelligence artificielle affective et les boucles affectives dans l'interaction humain-LLM, ainsi que les théories de l'innovation numérique (notamment le contrôle réversible). Ces travaux se traduisent par des manuscrits pour MISQ, Academy of Management Review et Communications of the ACM. J'ai récemment dirigé un projet de cartographie de plus de 250 000 familles de brevets en apprentissage automatique (2010–2025) afin d'identifier les domaines émergents liés à la détection des hallucinations, à la défense contre la prompt injection, à la sécurité des agents et à la fédéralisation de l'entraînement de LLM. Ce travail illustre ma manière de faire : croiser rigueur analytique, compréhension fine des organisations et préoccupations stratégiques.

Le programme de recherche que je propose s'articule autour de trois axes : (1) la gouvernance et la gestion des risques des systèmes d'IA dans les organisations ; (2) l'ingénierie, les opérations et la sécurité des systèmes d'IA (AIOps/MLOps) avec un focus sur la sécurité des LLM (prompt injection, hallucinations), les systèmes agentiques et l'IA affective ; et (3) la performance, l'impact et la durabilité de l'IA en contexte organisationnel. L'objectif est de produire à la fois des contributions théoriques (théories de l'innovation numérique, modèles de gouvernance, cadres d'évaluation de la sécurité des LLM, théories des boucles affectives) et des outils directement utiles aux organisations (référentiels de défense contre la prompt injection, patrons d'architecture pour systèmes RAG sécurisés, tableaux de bord, cas d'enseignement), en forte synergie avec le SIMQG, le Centre de recherche Createch sur les organisations intelligentes (CROI) et le Centre Laurent Beaudoin.

En enseignement, j'ai une expérience significative à tous les cycles dans des cours de systèmes d'information, d'analytique d'affaires, de transformation numérique et de gouvernance de l'IA. Ma philosophie est de partir de problèmes réels, d'outiller les étudiantes et étudiants pour les structurer

grâce aux données et aux modèles, puis de les amener à réfléchir aux implications organisationnelles et sociétales de leurs solutions. L'orientation pratique et partenariale de l'École de gestion correspond très bien à cette approche.

Enfin, je suis particulièrement attiré par la culture de collaboration interdisciplinaire et par l'accent mis sur la formation de leaders capables de diriger la transformation numérique s'appuyant sur l'IA. Je serais heureux de m'investir pleinement dans les projets du SIMQG, du CROI et de l'École de gestion, en recherche, en enseignement et en service à la collectivité.

Je vous remercie de l'attention portée à ma candidature et me tiens à votre disposition pour toute information complémentaire.

Veuillez agréer, Madame, Monsieur, l'expression de mes salutations distinguées.

[Signature]

Carlos Denner dos Santos [Courriel] [Téléphone]

Programme de recherche

Programme de recherche Gestion, gouvernance et ingénierie des systèmes d'intelligence artificielle

Depuis plus de vingt ans, je travaille à l'intersection des systèmes d'information, des données et de la gestion. Avec l'essor de l'IA, et en particulier des modèles de type LLM et des systèmes agentiques, la question centrale pour les organisations n'est plus "que peut-on faire avec l'IA ?", mais "comment la gouverner, l'exploiter et la sécuriser de façon responsable, performante et durable ?".

Le programme de recherche que je propose à l'Université de Sherbrooke vise précisément ce point. Il se structure en trois axes complémentaires :

1. Gouvernance et gestion des risques de l'IA dans les organisations
2. Ingénierie, opérations et sécurité des systèmes d'IA (AIOps/MLOps), avec un focus sur les LLM et agents
3. Performance, impact et durabilité de l'IA en contexte organisationnel

Ces axes s'inscrivent directement dans les thématiques du poste (gouvernance et risques, AIOps/MLOps, sécurité et résilience, performance des systèmes d'IA) et dans la mission du Département des systèmes d'information et méthodes quantitatives de gestion (SIMQG), au cœur de la transformation numérique, de l'analytique et de la cybersécurité.

Je m'appuie sur trois blocs d'expérience : – un ancrage académique en systèmes d'information et gouvernance de l'IA, avec des publications dans des revues à comité de lecture (notamment Ethics and Information Technology et Government Information Quarterly) ; – une pratique terrain de projets d'IA et d'analytique dans les télécommunications, l'énergie, la santé et le secteur public ; – un travail récent de cartographie de plus de 250 000 familles de brevets en IA/ML (2010–2025) pour identifier les domaines émergents liés à la gouvernance, la sécurité et l'ingénierie de l'IA.

Axe 1 – Gouvernance et gestion des risques de l'IA

Dans l'article « Artificial Intelligence Regulation : a framework for governance » (Ethics and Information Technology, 2021), coécrit avec des collègues, nous proposons un cadre intégrateur pour la régulation de l'IA à l'échelle des politiques publiques. Dans « Artificial intelligence governance : Understanding how public organizations implement it » (Government Information Quarterly, 2025), nous analysons comment 28 organisations publiques sur cinq continents traduisent (ou non) ces principes dans leurs pratiques.

Le premier axe prolonge ces travaux au niveau organisationnel et des portefeuilles d'IA, avec un accent particulier sur les risques émergents liés aux LLM et aux systèmes agentiques. Les questions centrales sont :

– Comment les organisations structurent-elles la gouvernance de leurs systèmes d'IA (rôles, comités, politiques, processus de décision) ? – Quels modèles de gouvernance (centralisé, fédéré, par domaine d'affaires) sont les plus adaptés selon le secteur et la maturité numérique ? – Comment cartographier de manière exploitable les risques liés à l'IA (biais, non-conformité réglementaire, cyberrisques spécifiques aux LLM comme la prompt injection et les hallucinations, dépendance à des fournisseurs, perte de contrôle sur des agents autonomes, boucles affectives incontrôlées dans les systèmes d'IA conversationnelle) pour aider les gestionnaires à prioriser ? – Comment les organisations peuvent-elles implémenter un « contrôle réversible » permettant d'ajuster dynamiquement leurs systèmes d'IA tout en maintenant des capacités d'apprentissage organisationnel ?

Je prévois de combiner études de cas approfondies, entretiens, analyse de documents et enquêtes quantitatives. L'objectif est de produire : – des typologies de structures de gouvernance de l'IA ; – des cartes de risques et de contrôles associés ; – des guides pratiques et cas d'enseignement utilisables dans les programmes de l'École de gestion (dont le microprogramme de 3e cycle en gestion stratégique de l'IA et l'École d'été en gestion stratégique de l'IA).

Axe 2 – Ingénierie, opérations et sécurité des systèmes d'IA (AIOps/MLOps)

Le deuxième axe se concentre sur le cycle de vie des systèmes d'IA : données, modèles, déploiement, exploitation, monitoring, sécurité. C'est le cœur de l'AIOps/MLOps, là où les enjeux de gouvernance se matérialisent vraiment. Mes travaux récents portent spécifiquement sur la sécurité des LLM et les dimensions affectives de l'interaction humain-IA.

Les questions de recherche incluent :

– Comment concevoir des pipelines AIOps/MLOps "gouvernables", où traçabilité, contrôles d'accès, revues de risques, audits et mécanismes d'arrêt ("kill switches") font partie de l'architecture et des outils, plutôt que de rester dans des documents ? – Quels patrons d'architecture pour les systèmes basés sur des LLM (RAG, agents, chaînes d'outils) permettent de garder la main sur ce que le système peut faire, sur quelles données et avec quelles garanties ? – Comment intégrer la sécurité (prompt injection, hallucinations critiques, exfiltration de données, agents sur-permissionnés) dans les pratiques MLOps, au même titre que la performance et la disponibilité ? – Comment construire un "pare-feu LLM" (LLM Firewall) avec des défenses multi-phases contre la prompt injection, depuis l'analyse de brevets jusqu'aux garde-fous déployables côté entrée ? – Comment évaluer et atténuer les hallucinations dans les systèmes de génération augmentée par récupération (RAG) ? – Comment les boucles affectives émergent-elles dans l'interaction humain-LLM, et quelles implications ont-elles pour la conception des systèmes d'IA conversationnelle et l'analyse de sentiment ?

Sur ce volet, je m'appuie sur : – des projets industriels concrets (par exemple un système d'optimisation énergétique pour un grand opérateur télécom, un système de recommandation IA pour l'application Jooay favorisant l'inclusion numérique des enfants et jeunes en situation de handicap, des projets de transformation numérique dans le secteur public) ; – un travail systématique sur les brevets en IA/ML montrant que la détection d'hallucinations, la défense contre la prompt injection, la sécurité des agents et la fédéralisation de l'entraînement des LLM sont des domaines émergents où l'activité reste faible au regard des enjeux ; – des manuscrits en préparation pour des revues de premier plan : « Building an LLM Firewall » (Communications of the ACM), « Evaluating and Mitigating Hallucinations in RAG » (cadre expérimental), « LLMs, Sentiment Analysis, and Algorithmic Feelings » (Academy of Management Review), et « Reversible Control as a Digital Innovation Theory » (MISQ Theory & Review).

Je privilégierai des approches de design science et de recherche orientée artefact : conception et évaluation de prototypes de pipelines MLOps intégrant des points de contrôle de gouvernance, d'"agents firewall" encadrant ce qu'un agent LLM peut faire, de systèmes RAG avec atténuation des hallucinations, et de tableaux de bord de risques et de performance opérationnelle. Ces artefacts seront développés et évalués avec des organisations partenaires (secteur public, télécom, énergie, santé), en visant à la fois des contributions scientifiques (théories, modèles, taxonomies, cadres d'évaluation) et des livrables directement utilisables.

Axe 3 – Performance, impact et durabilité de l'IA

Le troisième axe répond à une question que je rencontre régulièrement : l'IA crée-t-elle réellement de la valeur, pour qui et à quel coût ? Il s'agit ici de passer des promesses et des preuves de concept à une évaluation rigoureuse de la performance, de l'impact et de la durabilité des systèmes d'IA.

Les questions de recherche sont, par exemple :

– Comment définir et mesurer la performance de projets d'IA au-delà des métriques techniques (précision, F1) : valeur économique, qualité de service, impact sur les processus, équité, effets sur le travail humain ? – Quels facteurs distinguent les projets qui passent à l'échelle de ceux qui restent au stade de pilote (alignement stratégique, maturité des données, structures de gouvernance, capacités AIOps/MLOps, acceptation des utilisateurs) ? – Comment intégrer la durabilité (environnementale, économique, sociale) dans la priorisation et l'évaluation des portefeuilles d'IA ?

Je prévois des études de cas longitudinales (avant / après, quasi-expériences, analyses de séries

temporelles) combinées à la construction de tableaux de bord de performance co-conçus avec des partenaires. Les résultats nourriront directement les formations de l'École de gestion destinées aux gestionnaires et aux professionnelles et professionnels en exercice.

Approche, environnement et formation

Globalement, mon programme de recherche est pluridisciplinaire et multi-méthodes : ancré en systèmes d'information et en gestion de la technologie, combinant méthodes qualitatives, quantitatives et de design, et toujours connecté à des organisations réelles.

Le SIMQG, au cœur de la transformation numérique, de l'analytique et de la cybersécurité, est un environnement idéal pour développer ce programme en lien avec le Centre de recherche Createch sur les organisations intelligentes (CROI) et le Centre Laurent Beaudoin. Les projets envisagés se prêtent bien à des demandes de financement au CRSH, au FRQSC, à Mitacs et à des partenariats publics et privés.

Pour les étudiantes et étudiants de maîtrise, de doctorat et des microprogrammes de 3e cycle, ce programme offre un terrain riche pour des travaux qui combinent analyse rigoureuse, développement d'outils ou de méthodes, et impact concret dans les organisations. À horizon cinq ans, l'objectif est que l'École de gestion de l'Université de Sherbrooke soit reconnue comme un pôle de référence pour la gestion stratégique, la gouvernance et l'ingénierie des systèmes d'IA, au Québec et à l'international.

Énoncé d'enseignement

Énoncé d'enseignement Carlos Denner dos Santos

1. Philosophie d'enseignement

Je me vois d'abord comme quelqu'un qui aide les étudiantes et étudiants à utiliser les données, les systèmes d'information et l'intelligence artificielle pour résoudre des problèmes réels. Tout le reste – concepts, méthodes, outils – est au service de cet objectif.

Ma philosophie d'enseignement repose sur quatre principes :

1) Ancrer l'apprentissage dans des situations authentiques
   Je pars le plus souvent de problèmes ou de situations réels : un projet d'IA mal cadré, une organisation publique qui peine à gouverner ses données, une startup qui doit choisir un modèle d'affaires, ou un système de recommandation à améliorer. Cette approche par problèmes m'a guidé autant dans mes cours de systèmes d'information, de gouvernance de l'IA, d'analytique que dans mes activités de supervision.
2) Combiner rigueur conceptuelle et "learning by doing"
   Je tiens beaucoup à l'articulation entre théorie et pratique. Les étudiantes et étudiants doivent comprendre les modèles et cadres conceptuels, mais aussi les mettre en œuvre, les tester, les adapter. Dans mes cours, un concept important est presque toujours associé à un exercice appliqué : une étude de cas, une mini-enquête, une analyse de données, une simulation ou un prototype de système. Cette logique a également inspiré mes publications sur l'usage d'exercices en classe et de projets itératifs pour améliorer les cours.
3) Cultiver l'autonomie réflexive
   Je considère l'université comme un lieu où l'on apprend à poser de bonnes questions, pas seulement à produire de bonnes réponses. Je pousse les étudiantes et étudiants à réfléchir à ce qu'ils font : pourquoi ce modèle plutôt qu'un autre ? Quelles hypothèses implicites se cachent derrière cet indicateur ? Quelles sont les implications éthiques et organisationnelles de cette solution ? Cette réflexivité est centrale pour des sujets comme la gouvernance de l'IA, les systèmes intelligents et l'analytique d'affaires.
4) Construire un climat inclusif et exigeant
   Je cherche à créer un environnement où les étudiantes et étudiants se sentent à la fois en sécurité pour poser des questions et mis au défi de sortir de leur zone de confort. Cela implique des attentes claires, un feedback régulier et transparent, et une attention particulière aux différentes trajectoires (étudiantes et étudiants en emploi, internationaux, en reconversion, etc.). Ayant enseigné dans des contextes très divers (Brésil, Canada, programmes en présentiel et en ligne), je suis particulièrement sensible à ces écarts.

2. Expérience d'enseignement et champs de cours

Au fil de ma carrière, j'ai enseigné à tous les cycles dans des programmes de gestion, de systèmes d'information, de data science et d'informatique appliquée. Les cours que j'ai le plus souvent donnés ou co-conçus se situent dans les domaines suivants :

– Systèmes d'information de gestion et transformation numérique
– Gouvernance des TI et de l'IA, gestion des risques technologiques, cybersécurité organisationnelle et sécurité des systèmes LLM
– Analytique d'affaires, science des données et intelligence artificielle affective pour gestionnaires
– Programmation (Python), bases de données et systèmes RAG pour la gestion
– Entrepreneuriat, innovation numérique et théories de l'innovation (contrôle réversible, capacités de déploiement des connaissances)

Cette expérience se complète par des contributions à des ouvrages de cas et à des publications orientées pédagogie (exercices en classe, amélioration continue de cours, recours à des projets réels pour développer des compétences en entrepreneuriat et en génie logiciel), ce qui a renforcé ma conviction que l'enseignement peut et doit être un objet de recherche et d'expérimentation systématique.

3. Approches pédagogiques concrètes

Concrètement, mes cours s'organisent autour de quelques dispositifs clés.

a) Études de cas et projets appliqués
Je fais un usage intensif de cas – parfois tirés de la littérature, parfois construits à partir de mes propres expériences de recherche et de consultation – pour amener les étudiantes et étudiants à analyser des situations réalistes, à prendre position et à proposer des plans d'action. En parallèle, chaque fois que possible, je conçois des projets de session où les équipes travaillent sur un problème réel (organisation partenaire, données ouvertes, ou problématique issue de leurs milieux professionnels), avec des livrables intermédiaires et une restitution finale structurée.

b) Laboratoires et "studios" de données et d'IA
Pour les cours impliquant la programmation, l'analytique ou l'IA, je privilégie des séances de laboratoire où les étudiantes et étudiants manipulent eux-mêmes les données, les modèles et les outils (notebooks, plateformes analytiques, environnements de type MLOps/LLMOps simplifiés). L'idée est de passer rapidement du "voir faire" au "faire soi-même", en alternant mini-exposés, démonstrations et travail en autonomie guidée.

c) Intégration explicite des enjeux de gouvernance, d'éthique et de sécurité
Dans les cours sur l'IA, les systèmes d'information et l'analytique, j'intègre de manière systématique les questions de gouvernance, d'éthique, de biais, de transparence, de responsabilité et de sécurité des LLM. Concrètement, cela passe par des segments d'analyse de politiques (réglementations, lignes directrices), des discussions structurées sur des incidents réels (attaques par prompt injection, hallucinations critiques, boucles affectives problématiques), et des exercices où les étudiantes et étudiants doivent proposer des mécanismes de contrôle, des indicateurs, des défenses contre les vulnérabilités des LLM ou des structures de gouvernance adaptés.

d) Usage réfléchi des outils numériques et des LLM
Je considère les LLM et outils d'IA générative comme des objets à la fois d'étude et de pratique. Je les utilise comme "co-pilotes" dans certaines activités (par exemple pour générer des idées, explorer des options de modélisation, comparer des approches), tout en expliquant clairement leurs limites (hallucinations, biais, dépendance au contexte, risques de sécurité). Cela permet de développer chez les étudiantes et étudiants une posture de maîtrise critique vis-à-vis de ces outils, plutôt qu'une adoption naïve ou un rejet de principe.

4. Évaluation et feedback

L'évaluation de l'apprentissage est, pour moi, d'abord un problème de conception : que voulons-nous vraiment mesurer, et comment le rendre transparent pour les étudiantes et étudiants ?

Dans la plupart de mes cours, je combine :

– des travaux d'équipe (analyses de cas, projets de session, prototypes de solutions, rapports de recommandation) ;

– des évaluations individuelles (quiz courts pour valider la compréhension des concepts, mini-essais, journaux de bord réflexifs, examens centrés sur la résolution de problèmes) ;

– et une part de coévaluation ou d'auto-évaluation structurée, notamment dans les projets où la contribution individuelle peut varier.

Je consacre du temps en classe à expliciter les critères d'évaluation (rubriques), à commenter les premières livraisons et à ajuster si nécessaire pour éviter les malentendus. Je considère le feedback comme une partie intégrante du processus d'apprentissage, pas comme une simple justification de la note.

5. Encadrement et développement des étudiantes et étudiants

En plus des cours formels, j'ai encadré de nombreux travaux de fin d'études, mémoires de maîtrise et projets d'amélioration en organisation, notamment dans les domaines de la gouvernance des TI, de la transparence publique, de la gestion des incidents de sécurité et de la transformation numérique. Ces encadrements se caractérisent par :

– un cadrage initial solide de la question de recherche ou du problème de gestion ;
– un accompagnement méthodologique (choix de méthodes, collecte et analyse de données, validation) ;
– et une exigence de retombées réelles pour l'organisation partenaire (recommandations actionnables, instruments ou prototypes, cadres d'évaluation).

À l'Université de Sherbrooke, je souhaite contribuer activement à la supervision d'étudiantes et d'étudiants à la maîtrise et au doctorat, ainsi qu'aux projets du microprogramme de 3e cycle en gestion stratégique de l'IA. Mon objectif est de former des personnes capables de parler aussi bien le langage des dirigeants que celui des spécialistes techniques – un profil clé pour la gestion de l'IA.

6. Contribution à l'innovation pédagogique à l'Université de Sherbrooke

L'environnement de l'École de gestion et du SIMQG, fortement connecté aux milieux de pratique et à l'innovation pédagogique, est particulièrement propice pour prolonger ce travail. Concrètement, je vois plusieurs façons de contribuer :

– co-développer ou actualiser des cours sur la gestion stratégique de l'IA, la gouvernance des systèmes d'IA, l'analytique d'affaires et la transformation numérique, en les ancrant dans des cas et projets issus de mes recherches et collaborations ;
– participer à des initiatives d'innovation pédagogique (par exemple, développement de jeux sérieux ou de simulations autour de la gouvernance de l'IA, utilisation de plateformes d'analytique et de LLM en contexte de cours) ;
– contribuer à la création de ressources pédagogiques bilingues (français / anglais) sur la gouvernance, la gestion des risques et l'ingénierie des systèmes d'IA.

En résumé, mon approche de l'enseignement est profondément liée à ma recherche et à mes expériences sur le terrain : je cherche à outiller les étudiantes et étudiants pour qu'ils puissent concevoir, gouverner et critiquer des systèmes d'IA et de données dans des organisations réelles, avec à la fois rigueur conceptuelle et sens pratique. Je serais heureux de poursuivre et d'amplifier cette démarche au sein de l'École de gestion de l'Université de Sherbrooke.

Three Key Publications

Publication 1 : Artificial intelligence governance

Government Information Quarterly 42 (2025) 102003

Contents lists available at ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

Artificial intelligence governance: Understanding how public organizations implement it

Patricia Gomes Rêgo de Almeida [a,b,*], Carlos Denner dos Santos Júnior [a]

[a] University of Brasilia, Department of Business Management-PPGA, Campus Darcy Ribeiro, Prédio da FACE, Asa Norte, Brasília, DF 70910-900, Brazil
[b] Chamber of Deputies of Brazil, Palácio do Congresso Nacional - Praça dos Três Poderes, Brasília, DF 70160-900, Brazil

ARTICLE INFO

Keywords:
AI governance
AI regulation
Responsible AI
AI ethics
AI in public sector
AI in government
Fuzzy QCA

ABSTRACT

While observing the race for Artificial Intelligence (AI) regulation and global governance, public organizations are faced with the need to structure themselves so that their AI systems consider ethical principles. This research aimed to investigate how public organizations have incorporated the guidelines presented by academia, legislation, and international standards into their governance, management, and AI system development processes, focusing on ethical principles. Propositions were elaborated on the processes and practices recommended by literature specialized in AI governance. This entailed a comprehensive search that reached out to 28 public organizations across five continents that have AI systems in operation. Through an exploratory and descriptive aim, based on a qualitative and quantitative approach, the empirical analysis was carried out by means of proposition analysis using the Qualitative Comparative Analysis (QCA) method in crisp-set and fuzzy modes, based on questionnaire responses, combined with an interview and document content analysis. The analyses identified how processes and practices, across multiple layers and directed at the application of ethical principles in AI system production, have been combined and internalized in those public institutions. Organizations that trained decision-makers, AI system developers, and users showed a more advanced stage of AI governance; on the other hand, low scores were found on actions towards AI governance when those professionals did not receive any training. The results also revealed how governments can boost AI governance in public organizations by designing AI strategy, AI policy, AI ethical principles and publishing standards for that purpose to government agencies. The results also ground the design of the AIGov4Gov framework for public organizations to implement their own AI governance.

1. Introduction

After being coined "Artificial Intelligence" in 1956 (Cerka et al., 2017), a variety of research on AI has been developed, initially as knowledge-driven, later as data-driven, or combining them. The growing scope of AI in society has been observed through the combination of this technology with other emerging ones, generating the expression AI Plus (AI+) (Shao et al., 2022), boosting productivity (Mezgár & Váncza, 2022) and its influence on social transformation (Boyd & Holton, 2018).

The benefits offered by AI have reached the Government (Alhosani & Alhashmi, 2024). While Artificial Intelligence (AI) is seen as an enabler of digital transformation for organizations (Holmström, 2022; Kitsios & Kamariotou, 2021), in the public sector, governments' development strategies coincide with their AI strategies (Wirtz et al., 2018). However, at the same time, concerns grow about ethical impacts on AI-dependent decisions when ethical principles are not considered (Ashok et al., 2022; Bonsón et al., 2021; Hopster, 2021; Kazim & Koshiyama, 2021; Stahl et al., 2022; Wirtz et al., 2022). Immersed in this scenario, the movement for a responsible AI (Eke et al., 2023) using AI regulation and governance has involved governments, academia, and international standardization bodies (Carter, 2020; Gianni et al., 2022; Gutierrez & Marchant, 2021; IEEE, 2019, 2020, 2021a, 2021b, 2021c, 2021d, 2022; ISO, 2021a, 2021b, 2021c, 2022a, 2022b; OECD, 2022c).

Even though there is a significant number of theoretical essays (De Almeida et al., 2021), AI governance is still an underdeveloped area of research (Morley et al., 2020; Taeihagh, 2021), requiring a greater understanding of how organizations have interpreted and incorporated

* Corresponding author at: University of Brasilia, Department of Business Management, Campus Darcy Ribeiro, Prédio da FACE, Asa Norte, Brasília, DF 70910-900, Brazil.
E-mail addresses: patricia.rego.almeida@gmail.com (P.G.R. de Almeida), carlosdenner@unb.br (C.D. dos Santos Júnior).

https://doi.org/10.1016/j.giq.2024.102003
Received 11 June 2023; Received in revised form 25 September 2024; Accepted 20 December 2024
Available online 16 January 2025

ethical principles into their practices, processes and structures when producing AI systems (Mäntymäki et al., 2022; Mikalef, Conboy, et al., 2022). In systematic research on AI in public governance, Zuiderwijk et al. (2021) identified research gaps that adopt multiple methods, combining exploratory empirical research with qualitative-quantitative analyses, in order to delve deeper into practices used by AI governance in the public sector. Considering the important role that government bodies play in AI regulation and governance (Cihon et al., 2020; Stix, 2021), the potential benefits and risks that AI can bring to society when it supports the public sector (Ahn & Chen, 2022; Ojo et al., 2019; Sharma et al., 2020), and the challenge of avoiding loss of confidence in AI-supported government decisions (Zuiderwijk et al., 2021), it becomes crucial to understand how public organizations are following the academia, legislation, and standard recommendations concerning ethical principles in the use and development of AI systems.

2. AI governance

An AI that considers ethical principles brings new obligations to organizations (Hickman & Petrin, 2021; Roorda, 2021; Smuha, 2021). However, traditional governance processes and structures are not sufficient for the challenges posed by AI governance (Taeihagh, 2021). In public institutions, the challenge is greater because, in general, citizens do not choose AI products but are obligated to consume them as they are embedded in public services (Zuiderwijk et al., 2021).

In the context of AI governance in public organizations, ethical principles are applied to maximize the benefits of AI and minimize its risks (Rose et al., 2018; Vial, 2019). Considering the focus on impacts generated in society as a premise for AI governance (Djeffal, 2018), researchers point out the need to integrate the Stakeholder Theory with the Social Contract Theory (Bonsón et al., 2021; Wright & Schultz, 2018) to obtain society's perceptions of the values involved in decisions made by AI systems (Hickman & Petrin, 2021; Rahwan, 2017). In the corporate context, subordinated to IT governance (IT Governance Institute, 2003), an effective AI governance requires a multilayered model, the upper layer of which includes mechanisms for capturing government regulatory requirements and legislation, translating them into the organizational context through internal regulations that establish ethical principles and conditions for their application through processes and practices (Mäntymäki et al., 2022), as illustrated by the conceptual research model in Fig. 1a, whose details are found in the subsections 2.1 up to 2.6.

Considering many initiatives to regulate AI through legislation, government policies, and international standards (Fjeld et al., 2020; Gutierrez & Marchant, 2021; OECD, 2022c), it is expected that public organizations implement their own AI governance aligned with such regulation initiatives.

2.1. AI governance actions at the strategic level of public organizations

The relationship among the different scopes of governance, defined by Mäntymäki et al. (2022), establishes that corporate governance contains IT governance, which, in turn, contains AI governance. Consequently, AI governance inherits characteristics from IT governance.

IT governance relates to IT decision-making at the board of directors and executive management, which involves: creating an organizational structure (unit, committee, board), elaborating a strategy to effectively address the organization's needs through AI (Herremans, 2021), and implementing processes that support the board's decisions (Aasi et al., 2014). To formalize IT governance, organizations establish policies (Mäntymäki et al., 2022). Thus, similar to those actions that demonstrate the existence of IT governance at a higher-level decision-making board (Aasi et al., 2014), high-level decision-making actions for AI governance were considered in this research.

Indeed, in Papagiannidis et al.'s (2023) research, the higher-level decision-makers highlighted the importance of an AI strategy to manage the corporate needs that AI systems will address, as well as the AI governance process in their organization. In the same sense, Agarwal's (2023) research points out that the existence of an AI governance structure at a high level of the organization is crucial for setting the organization's strategic direction in AI-related initiatives. Complementing them, Sigfrids et al. (2023) emphasize that, in the public sector, AI policies should consider AI ethical principles, giving special

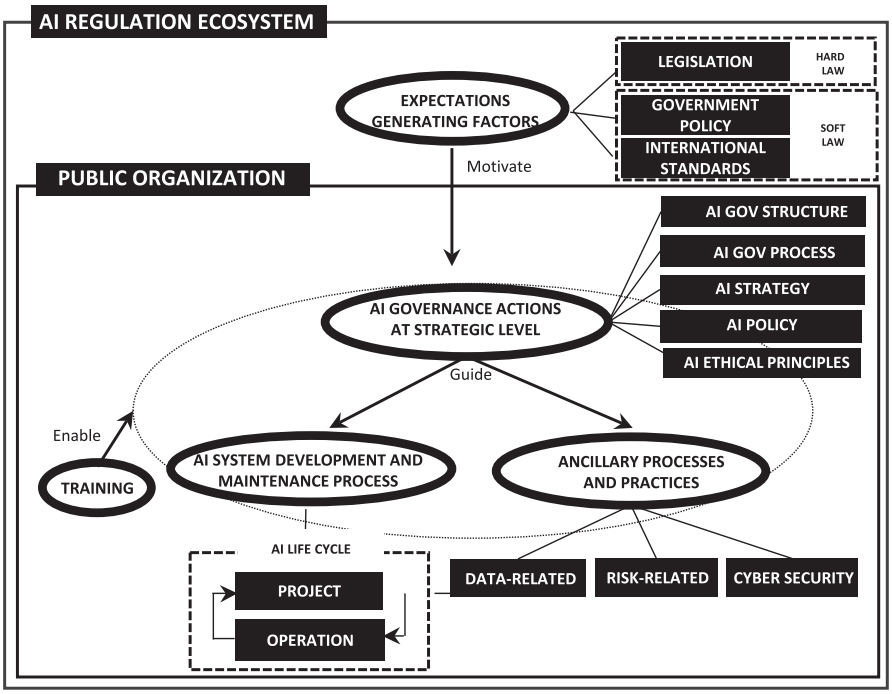


Fig. 1. a: Conceptual research model for AI governance in public organizations. (Source: Self-elaboration.)

attention to a wider socio-technical and political approach instead of merely respecting moral minimums. It requires considering values and gaining citizens' trust. Such an approach imposes establishing an AI ethical principles code at a high level of the organization. Thus, for the context of this research, "creating an AI governance structure," "elaborating an AI strategy," "establishing an AI policy," "implementing an AI governance process," and "establishing an AI ethical principles code" were at the higher level of the conceptual research model and were called "AI governance actions at a strategic level."

Guided by the AI governance actions at a strategic level, ancillary mechanisms are created to establish data-related processes (Janssen et al., 2020; Rhahla et al., 2021), a cybersecurity-related process (Breier et al., 2020; Xue et al., 2020), risk-related processes and practices (NIST, 2022; Wirtz et al., 2022) (Fig. 1a – Ancillary processes and practices) and AI system development-related processes that address AI ethical issues (De Silva & Alahakoon, 2022; Laato et al., 2022) (Fig. 1a – AI system development and maintenance process). Based on the above, the following **Proposition 1** is conjectured: "data-related," "risk-related," "cyber-security," and "AI system development" practices must follow guidelines established at the strategic level of AI governance. Considering that each dimension addressed in Proposition 1 can be decomposed into other actions, for greater accuracy of this study, derived propositions were created for each group of processes and practices.

2.2. Data-related processes and practices (Fig. 1a – Data-related processes and practices)

Given AI's reliance on data, a link is established between data quality and the outcome of AI systems (Dwivedi et al., 2021; Kuziemski & Misuraca, 2020). For this reason, focusing on reliable AI systems and data governance becomes crucial (Haneem et al., 2019; Vining et al., 2022).

Sometimes intersecting with AI governance (Alshahrani et al., 2021; Andrews, 2018; Dwivedi et al., 2021; Mäntymäki et al., 2022; Medaglia et al., 2021; Özdemir & Hekim, 2018), a data governance process specifies responsibilities over decisions made about the organization's data, as well as formalizes policies and standards (Abraham et al., 2019; Carretero et al., 2017; Vilminko-Heikkinen & Pekkola, 2019), which requires a great deal of stakeholders' negotiation skills to obtain consensus (Benfeldt et al., 2020; Calzada & Almirall, 2020; Micheli et al., 2020; Ruijer, 2021). Following data governance policies (Carretero et al., 2017), processes are defined for managing data quality (Haneem et al., 2019; Khatri, 2016) and personal data protection (Janssen et al., 2020). For those reasons, **Proposition 1 A** is based on the decomposition of the "data-related" construct into three processes: data governance process, data quality management process, and personal data protection management process, which must follow guidelines established at the strategic level of AI governance.

2.3. Risk-related processes and practices (Fig. 1a – risk-related processes and practices)

Proposition 1B comprises actions created to mitigate the risks posed by AI systems (Vetrò et al., 2021; Wirtz et al., 2022). However, traditional risk management processes, supported by quantitative methods (Chen & Deng, 2022; Duijm, 2015), have been criticized, requiring complementary approaches that integrate visions and thus also include a qualitative approach (Aiken, 2021; Fernandes et al., 2021; Gerkensmeier & Ratter, 2018; ISO, 2022a; ISO, 2022b).

Such integrated vision starts with the definition of stakeholders that are impacted, whether directly or indirectly, by the AI system (NIST, 2022; Wirtz et al., 2022). Therefore, identifying stakeholders in the whole AI lifecycle is a requirement for risk management (Wright & Schultz, 2018). In the same direction, audit processes for AI systems are also risk-oriented (De Oliveira, 2019; Erlina et al., 2020), associating them with stakeholders (Zicari et al., 2021). In addition, over time, changes in environmental variables can alter the context for which the AI system was designed, causing behaviors that differ from the desired results. This situation can be avoided by monitoring changes in the environment (rules, social trends, etc.) and feeding the risk management process (González et al., 2020). Thus, **Proposition 1B** was formulated as follows: risk management processes, audit processes, practices for identifying stakeholders, and practices for monitoring changes in the environment, all of which must follow guidelines established at the strategic level of AI governance.

2.4. Cybersecurity process (Fig. 1a – Cybersecurity)

Integrated with risk management (Breier et al., 2020; European Union Agency for Cyber Security, 2022), a cybersecurity management process is applied to prevent cyberattacks explicitly designed to exploit vulnerabilities in AI algorithms (Chen et al., 2019; Eggers & Sample, 2020; Gu et al., 2019; McGraw et al., 2020; Xue et al., 2020). To face those threats, organizations implement a security management process (European Union Agency for Cyber Security, 2021) that addresses the entire AI system lifecycle (Jing et al., 2021). Therefore, **Proposition 1C** was thus formulated as follows: the AI system security management process must follow guidelines established at the strategic level of AI governance.

2.5. AI system development process (Fig. 1a – AI system development and maintenance process)

Seeking to encompass the entire AI lifecycle, **Proposition 1D** was formulated as follows: practices aimed at ethical principles for AI system development and maintenance processes, both as a project and as an operational product, must follow guidelines established at the strategic level of AI governance. To investigate practices at each phase, Proposition 1D was further broken down into 1D1 and 1D2 to analyze the project phase and the operation phase, respectively.

At the project's starting point (Fig. 2 – Project), the translation of ethical principles into rules on the behavior of the system is deepened (Dennis et al., 2016; IEEE, 2021a), focusing on the definition of groups and attributes to be protected (González et al., 2020; ISO, 2021a; Rajkomar et al., 2018). Rules and ethical dilemmas are identified and analyzed (Anderson & Anderson, 2018; Awad et al., 2020; Bench-Capon & Modgil, 2017; Bonnemains et al., 2018; Locher & Bolander, 2019; Ma et al., 2018; Schrader & Ghosh, 2018; Zicari et al., 2021). Data extraction and preparation tasks demand attention to understand their characteristics and quality, preparing them for pre-processing (De Silva & Alahakoon, 2022). Techniques are applied to identify and minimize data biases (Ashokan & Haas, 2021; Baeza-Yates, 2018; ISO, 2021a; Leavy et al., 2020; Lin et al., 2021; Ntoutsi et al., 2020; Oneto & Chiappa, 2020; Roselli et al., 2019; Silberg & Manyika, 2019), while adjustments are made in the data sample (González et al., 2020). The building and validation of models involve algorithmic research, which requires decisions that also need to be free of bias (Abdollahi & Nasraoui, 2018; Ashokan & Haas, 2021; De Silva & Alahakoon, 2022; Makhlouf et al., 2021). Transparency practices are required to explain the system results (Adadi & Berrada, 2018; Arrieta et al., 2020; Das, 2020; Dazeley et al., 2021; Kale et al., 2022; Phillips et al., 2021). Focusing on the AI system project phase, **Proposition 1D1** was formulated: practices for representing rules and ethical dilemmas, practices for minimizing biases, and practices for providing transparency in the AI system development process must follow guidelines at the strategic level of AI governance.

The sensitivity to changes in the context for which the AI system was created, combined with the fact that AI models are less complex than social realities (Strauß, 2021), impose continuous monitoring after the AI system is in operation (Laato et al., 2022), which implies: a) automatic performance monitoring in charge of the IT staff (De Silva & Alahakoon, 2022; Fjeld et al., 2020; González et al., 2020), b) human oversight of the AI system behavior, usually by someone delegated by

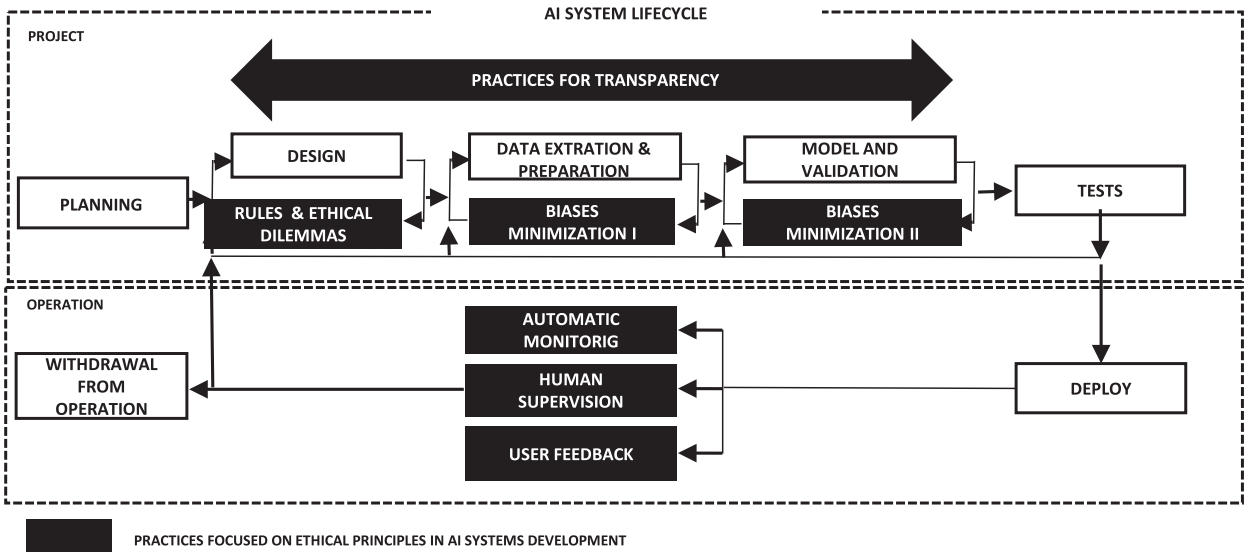


Fig. 2. Conceptual research model for the entire AI system lifecycle.
(Source: Self-elaboration)

the domains' decision-maker (Dignum, 2019; Fjeld et al., 2020; Hickman & Petrin, 2021; Strauß, 2021; Zicari et al., 2021), and c) continuous user feedback, which is often implemented through a feature in the AI system that asks its users' satisfaction with the system outputs and requests tips to refine its performance (Rahwan et al., 2019; Wright & Schultz, 2018) (Fig. 2 – Operation). The combination of those actions motivates the evolution of AI systems in a continuous loop until their withdrawal from operation (De Silva & Alahakoon, 2022; ISO, 2022a; Laato et al., 2022) (Fig. 2). Therefore, with the practices described in the AI system operation phase in mind, **Proposition 1D2** was formulated: practices for automatic monitoring, human oversight, and user feedback must follow guidelines established at the strategic level of AI governance.

2.6. *Training people (Fig. 1 – Training)*

Mikalef, Lemmer, et al. (2022) confirm that the decision-makers' perceptions regarding the potential AI value are drivers for AI adoption in public organizations. In the same sense, educating stakeholders on AI ethics becomes crucial for an effective AI ethical principles implementation in AI system production (Zhou & Chen, 2023), which includes a culture with fewer biases in organizations (Awad et al., 2020; Ma et al., 2018) as well as knowledge about practices required to AI governance (Ligot, 2024). With such purposes, training key stakeholders (decision-makers, developers, system users, and auditors) on AI ethical principles is considered an enabler for AI governance implementation in organizations (Calzada & Almirall, 2020; Herremans, 2021; Makarius et al., 2020; Micheli et al., 2020; Ruijer, 2021). Researchers point to an AI literacy program as a pivotal action to an effective and responsible AI adoption since it is a set of training and awareness actions that reaches staff beyond the IT team, which includes business decision-makers and system users. Decision-makers need it for a wide understanding of how AI can create value for their work processes and the requirements necessary for it (data quality, data protection, data governance, inclusive teams, for example). Users are also considered key stakeholders due to their role in data quality and data protection processes, and there is also the need for an awareness approach regarding ethical implications when AI is used inadequately (Pinski et al., 2024; Schüller, 2022).

Aiming to determine the enabling requirements of training people for AI governance practices, **Proposition 2** was elaborated: training stakeholders on data, the development of AI systems, and ethical principles applied to AI enables the implementation of AI governance in public organizations.

Considering that the complexity of internalizing AI governance in the organizations' processes and structures (Agarwal, 2023) can impact the "make-or-buy" decision on developing AI systems. Gräf et al. (2024) highlight that the opacity of AI systems and the lack of skilled human resources to deal with the whole AI system lifecycle are key factors for such decisions. In general, the decision to buy AI systems is an obstacle to accessing their codes, resulting in a transparency issue (Martin & Parmar, 2024). Adding such concerns to Mikalef, Lemmer, et al. (2022), Ahn and Chen's (2022) and Benfeldt et al.'s (2020) recommendations for training both the managerial and the technical spheres to deal with challenges such as reducing algorithmic opacity, it is possible to formulate **Proposition 3**: over time, in public organizations, training managers and AI system developers on AI ethics sparks interest in obtaining knowledge of AI system codes and contributes positively to AI governance.

3. Research methods and techniques

As an exploratory and descriptive study, the investigation was carried out through empirical research with a qualitative and quantitative approach to fill the gap identified by Mäntymäki et al. (2022) and Zuiderwijk et al. (2021) concerning knowledge of how organizations have interpreted and incorporated ethical principles in AI system production into their practices and processes, as especially demanded by Zuiderwijk et al. (2021) for using data-driven methods with exploratory and multiple approaches to deepen AI governance in the public sector.

3.1. *Sample selection and data collection strategy*

Since AI governance is a global need (Fjeld et al., 2020; OECD, 2022a), an attempt was made to build the sample including populations of public organizations belonging to any branches — Executive, Legislative, or Judiciary (Maluf, 1995) — from five continents. Considering the interest in investigating processes and practices, a search criterion was defined: a public organization should have at least one AI system in operation as part of its official portfolio. As a consequence, organizations that only had AI systems at a prototype stage or projects in an embryonic stage were not included in the sample.

The identification, sample selection, and data collection tasks were

performed over three months, consisting of several steps in parallel lines of research involving different actors and sources of information, as shown in Fig. 3 (MRE, 2022; OECD, 2022a); European Commission, 2018). Those communications took place through remote meetings, email, and social media until the contact information from researchers, government and AI system development and research centers, managers in charge of AI or digital transformation strategies, and AI use cases could be obtained. The path described in Fig. 3 culminated in 711 AI systems being developed or used by public organizations (European Commission, 2021b; FCT, 2021; Government of India, 2020; IPS-X, 2021; Misuraca & van Noordt, 2020; OECD/CAF, 2022; Tangi et al., 2022; WEF, 2020, 2020b). After removing redundancies and non-operational systems (prototypes or withdrawn from use), finetuning continued until the contact information for organizations that met the search criteria could be found. Upon invitation, 28 organizations effectively participated in the survey.

3.2. Data collection mechanisms

For the quantitative and qualitative analyses, primary data was used by means of an online questionnaire and semi-structured interviews (Appendix A - Annex 1a and 1b), both of which required basic knowledge of the decisions, processes, and practices related to AI system production. For this reason, the questionnaire was sent out only after the organization indicated a person in charge of the AI system portfolio. The interviews were conducted remotely and, in some cases, two or three individuals represented the organization.

Both the questionnaire and the interview script were submitted to a group of judges for assessment (PhD and Master researchers on ethical AI and data analysis) using methods indicated in the literature for each case (Fig. 4). For the questionnaire, the "Content Validation Coefficient" (CVC) (Aburachid & Greco, 2011; Hernández-Nieto, 2002; Silveira et al., 2018) was used with each question being evaluated on a scale from 1 to 5 to measure levels of clarity and relevance for the research (Appendix A - Annex 2 A). The interview script was evaluated using the "Validation for Qualitative Research Instruments" method (VALI-QUALI) (Torlig et al., 2022), an evolution of MRPQ (Torlig et al., 2019), taking the "Content" and "Semantics" dimensions into consideration. The content evaluation provided a score from 1 to 5 for each question in relation to the "alignment of each question with the research objective" and "adherence of the question to the investigated construct" attributes. The semantic analysis considered the "clarity" and "qualitative expectation of answer for each question" attributes (Appendix A - Annex 2). Pre-tests were carried out to the complete set (questionnaire and interview) using people with a similar profile to the target audience to confirm alignment with the research objectives (Manzini, 2004).

For each question regarding the existence of a practice or process implemented (or being implemented), four scores were considered according to standard answers (100 for "Yes, completely," 67 for "Yes, but only partially," 33 for "No, but a formal decision has been made to implement it," and 0 for "No, and no formal decision has been made to implement it." For the AI governance process and for practices aimed at transparency, adapted answers were made, once they were gathered from the interview (Appendix A - Annex 3).

3.3. Analysis strategy

The analysis of the primary data was carried out using a combination of Qualitative Comparative Analysis - QCA (Ragin, 2008; Rihoux & Ragin, 2008), and content analysis of the interviews and shared documents (Krippendorff, 2013; Saldaña, 2013).

3.3.1. QCA

The QCA is a qualitative research technique that also considers quantitative aspects for samples from 3 to 250 cases (Dias, 2011). Based on Set Theory and Boolean operations to establish logical relationships between sets, the QCA proposes to solve problems whose analysis requires causal inferences in case studies. The method seeks to show which combinations of conditions occurred in a scenario of an expected outcome (Rihoux & Ragin, 2008) to carry out comparative analyses, through associations between certain conditions and the outcome, instead of correlations (Korjani & Mendel, 2012; Ragin, 2008). A condition is required for a given result if the condition is always present when the outcome occurs. A condition is sufficient for a certain outcome

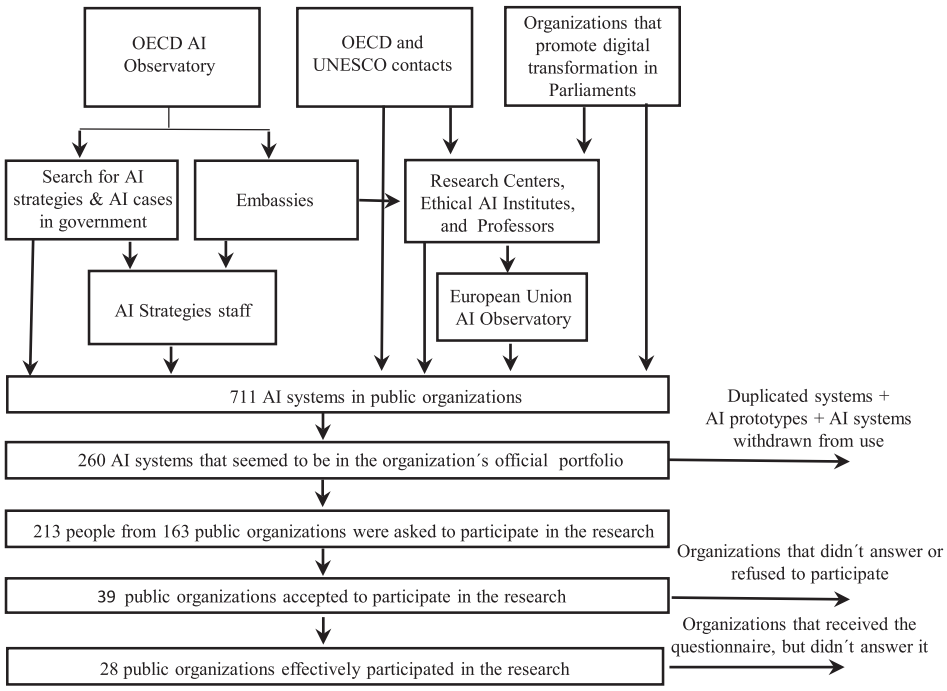


Fig. 3. Sample selection strategy.
(Source: Self elaboration)

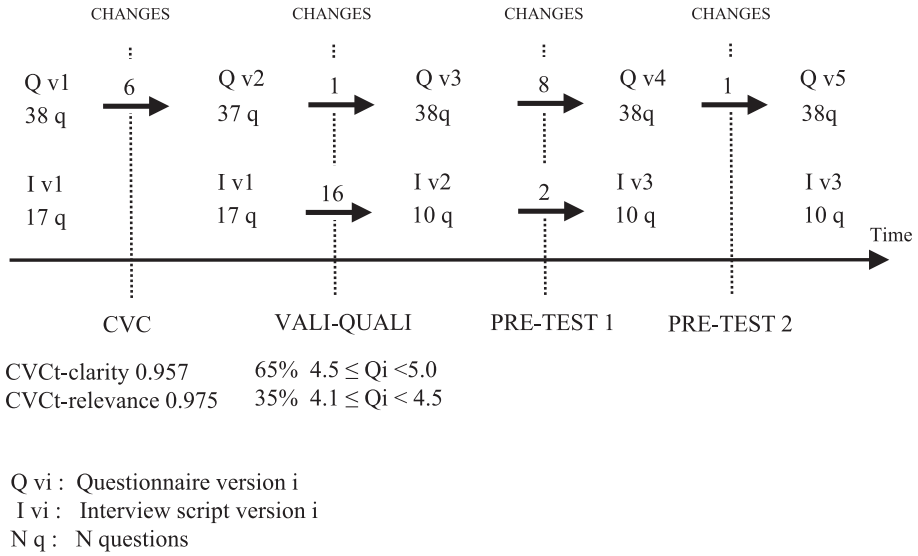


Fig. 4. Evolution of data gathering mechanisms.
(Source: Self-elaboration)

if this result always occurs when the condition is present (Rihoux & Ragin, 2008).

This research used both the crisp-set QCA (values 0 and 1, respectively, for the absence or presence of a relationship between the sets) and the fuzzy QCA, which offers more precision due to the use of a continuous set of values in the interval from 0 (complete absence of membership) to 1 (full membership) (Ragin, 2008; Rihoux & Ragin, 2008). In fuzzy analyses, A is a fuzzy subset of a fuzzy set B if given two fuzzy sets, A = {s1,s2,.., sn} and B = {g1,g2,.., gn}, and si, and gi, being i-th case scores in each set, and si ϵ [0;1] $\subset$ ℜ, gi ϵ [0;1] $\subset$ ℜ, $\forall$i; so, A $\subset$ B if si $\leq$ gi, $\forall$i. A calibration (Freitas & Neto, 2016; Meijerink & Bondarouk, 2018; Ragin, 2008) was made using the original values of the sets, turning them into fuzzy sets, whose values are distributed in the interval between 0 and 1, based on the presence level of the conditions in the outcome set. The main criteria for validating the fuzzy QCA is the consistency indicator to measure the relationship proximity between sets, indicating the degree to which the cases that share a combination of conditions agree with the outcome, with a value between 0 and 1. Consistency $(Xi \leq Yi) = \sum \frac{MIN(Xi,Yi)}{\sum Xi}, \forall i$; where X is the membership score in the causal combination, and Y is the membership score in the outcome (Betarelli-Júnior & Ferreira, 2018). Complementing the outcome interpretation, the coverage indicator offers the quantification of the empirical relevance of a causal combination in the causal combination set: Coverage $(Xi \leq Yi) = \sum \frac{MIN(Xi,Yi)}{\sum Yi}, \forall i$; where X is the membership score in the causal combination and Y is the membership score in the outcome (Rihoux & Ragin, 2008).

3.3.2. *Analysis plan*

A crisp-set QCA was applied for dichotomous variables, and the fuzzy QCA for variables with continuous values (Fig. 5) was associated with the research model constructs. The QCA was applied to analyze the propositions. As for the crisp-set QCA, the TOSMANA software version 1.6.1 (https://www.tosmana.net/) was used, while fsQCA version 3.1 (http://www.socsci.uci.edu/~cragin/fsQCA/software.shtml) was utilized for the fuzzy QCA. Another piece of information was extracted from the analysis of the interview content using MAXQDA 2022 software (https://www.maxqda.com/). Lastly, the union of the two analyses grounded the discussions to reach the conclusions that support the proposed framework.

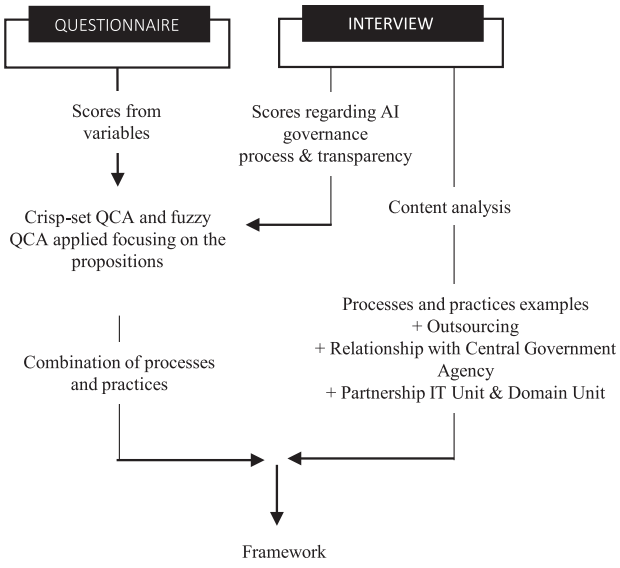


Fig. 5. Analysis plan.
(Source: Self-elaboration)

4. Results and discussions

The participants were people aware of the processes involved in the whole AI lifecycle. They were decision-makers and/or people appointed by them to participate. Table 1a displays the participants' profiles. The sample, resulting from the path taken as shown in Fig. 3, consists of 28 public organizations distributed across five continents, whose characteristics are listed in Tables 1b and 1c. The category of public organizations reveals their importance to citizens. The questionnaire asked the participant to provide information about the organization's five most frequently used AI systems (Appendix A - Annex 5).

The analyses that simultaneously involved dichotomous variables and continuous-value variables were performed by converting variables from continuous into binary and applying the crisp-set QCA (Proposition 3). For analyses in which all variables had continuous values (Propositions 1 A, 1B, 1C, 1D1, 1D2, and 2), the fuzzy QCA was applied with a fuzzy value calibration of 0.05 and 0.95, respectively, for each set's lowest and highest values, as also established by Navarro et al. (2016)

Table 1a
Interviewed profiles.

Organization Unit	N	%	Education	N	%	Position	N	%
Information Technology	17	60.7	PhD	9	32.1	Director/Manager	20	71.4
Innovation	2	7.14	Master	12	42.9	Data Scientist/IT Analyst	6	21.4
Data Science and Statistics*	3	10.7	MBA	4	14.3	Consultant	2	7.14
Corporate Governance	4	14.3	Undergraduate	3	10.7			
Others	2	7.14						

* Cases in which the data science unit is detached from the IT unit.

Tables 1b and 1c
Characteristics of the 28 public organizations in the sample used in the study.

Country	N	%	Coverage	N	%
Angola	1	3.57	National	24	85.71
Argentina	3	10.71	Group of countries	1	3.57
Australia	1	3.57	State/City	3	10.71
Brazil	4	14.29			
Canada	2	7.14	**Branch of Government**	**N**	**%**
Denmark	1	3.57	Executive	17	60.7
Estonia	2	7.14	Legislative	9	32.1
Finland	2	7.14	Judiciary	2	7.14
Germany	2	7.14			
Iceland	1	3.57	**No of Employees**	**N**	**%**
Italy	1	3.57	Up to 100	2	7.14
Japan	1	3.57	101 up to 500	3	10.71
Luxembourg	1	3.57	501 up to 1000	8	28.57
Norway	3	10.71	1001 up to 10,000	12	42.86
Sweden	1	3.57	More than 10,000	3	10.71
Switzerland	1	3.57			

Category	N	%
Parliament	8	28.57
Ministry of Finances\National Agency for Trade and Investment \National Business Authority \National Tax Agency\National Agency for Improving Business	6	21.42
National Agency for Unemployment Insurance Fund\Federal Office for Migration and Refugees\National Agency for Labour and Welfare	3	10.71
National Statistic and Research Institute\National Institute of Research and Technology	3	10.71
Federal Court/Federal Court Dept Specialized in Child Sexual Abuse Crimes	2	7.14
Hospital and Health Research Institute/National Agency for Auditing Businesses Producing Food	2	7.14
National Agency for Account Auditing	1	3.57
State Agency for Innovation	1	3.57
State Agency for Public Transport	1	3.57
Chief of Ministerial President Cabinet	1	3.57

AI systems production time	**N**	**%**
Less than 1 year	1	3.57
1–3 years	8	28.57
3–5 years	11	39.29
More than 5 years	8	28.57

and Codá et al. (2022). The descriptions of the variables are in Appendix A - Annex 6.

Regarding legislation, it is worth mentioning that 100 % of the sample comprises organizations subject to some law that protects personal data. In the sample's scope, only Denmark (Danish Government, 2020) approved a law on data ethics that also addresses AI systems development. Among the bills under discussion, the European AI Act (European Commission, 2021a; OECD, 2022b), despite being in an advanced discussion stage when the data were collected, was not yet approved as a law (European Parliament, 2022a; European Parliament, 2022b).

4.1. Analysis of proposition 1

Proposition 1 was analyzed through its decomposition into dimensions – data, risks, cyber security, and development – in Propositions 1 A, 1B, 1C, and 1D analyses, respectively. In those analyses, to represent AI governance actions at the strategic level of the organization, the following was considered (Appendix A - Annex 4): AI strategy; policy or recommendations directed at AI systems; guide to ethical principles that are applied to AI systems; AI governance processes; and the existence of a structure (unit, position, or board/council/committee to address AI governance). Those analyses carried out the calibration for turning the original values into fuzzy values while considering that the analyses focus on the conditions whose practices and processes have been implemented in any proportion (scoring options 67 or 100 – Appendix A - Annex 3). For those purposes, in all research analyses that used a fuzzy QCA, the crossing point was defined as 60.

4.1.1. Analysis of proposition 1 A

From the decomposition of the practices regarding data into processes for data governance (PDGOV), for data quality management (PDQUA), and for personal data protection management (PDPRO), an analysis of Proposition 1 A was carried out, considering, as an output set, the cases that were in a more advanced stage of implementing AI governance actions at a strategic level (SAIGOV(1)) (Appendix A - Annex 6). The process for managing personal data protection had the highest average (90.5, see Appendix A - Annex 4), which illustrates the most advanced stage of actions to comply with the personal data protection law to which they are subject.

The three preliminary combinations presented by the fuzzy QCA (Table 2) were not valid to the conclusion (combinations 1 and 2 had a consistency below the minimum value accepted by Ragin, 2008 and Scheider and Wagemann, 2012, and combination 3 was not conclusive). However, it should be noted that nineteen cases implemented the three processes – data governance, data quality management, and personal data protection management; and that, of the twelve highest scores for strategic actions towards AI governance, nine cases (75 %, see Appendix A - Annex 4) had implemented, at any stage, all three processes. Due to that, the "Necessary Condition" assessment test was applied to the combination of the three processes, which resulted in a consistency of 0.959862, revealing that high values would have been unlikely to occur in AI governance actions at a strategic level without the simultaneous existence of the three processes tested. Thus, the simultaneous implementation of the three processes is associated with organizations at an advanced stage in those AI governance actions at the strategic level defined in 2.1.

The positive impact when the three data processes are implemented aligns with the researchers' arguments that data governance is crucial for AI governance (Haneem et al., 2019; Vining et al., 2022) since data governance defines stakeholders' roles and responsibilities concerning the data (Sivarajah et al., 2017), which, in turn, requires managers' involvement in many deliberations (Benfeldt et al., 2020) to improve data quality process (Haneem et al., 2019; Rhahla et al., 2021; Vilminko-Heikkinen & Pekkola, 2019) and personal data protection management process. Thus, data governance guides data quality management processes and personal data protection processes (Labadie et al., 2020).

Table 2
Fuzzy QCA applied to proposition 1 A.

Proposition		Combinations	Cases	Results	Coverage and Consistency
1A	1	PDGOV * PDPRO	3,5,9,19,22,27,6,8,13,25	SAIGOV(1)	Cv: 0.825606
			4,20,24,2,7,11,15,16,23	SAIGOV(0)	Cs: 0.71781
	2	PDQUA * PDPRO	8,19,22,25,5,6,9,13,27	SAIGOV(1)	Cv: 0.733564
			20,23,2,4,7,21,15,16,18,28	SAIGOV(0)	Cs: 0.688312
	3	~ PDGOV* ~ PDQUA* ~ PDPRO	21	SAIGOV(1)	Cv: 0.215917
			1	SAIGOV(0)	Cs: 0.75
	"Necessary Conditions" Test				Cv:0.581795
	PDGOV * PDQUA * PDPRO			SAIGOV(1)	Cs: 0.959862

4.1.2. Analysis of proposition 1B

Proposition 1B was analyzed based on the decomposition of risk-related practices into a process for risk management (PRISM), a process for auditing AI systems (PAUDIT), a practice for identifying stakeholders (STAKEH), monitoring changes in the environment and social trends (ENVIRO), considering cases that were at a more advanced stage in the implementation of AI governance actions at a strategic level (SAIGOV(1)) (Appendix A - Annex 6). Of the three combinations that were calculated (Table 3), only combination 2 showed a proportion favorable to advanced-stage cases in AI governance actions at a strategic level, which presented 85.71 % of the cases (cases 3, 8, 9, 13, 22, 27) that have implemented, at any stage, a risk management process, practices for identifying stakeholders, and practices for monitoring environmental changes and social trends (consistency = 0.941645). Therefore, those processes and practices can be considered to have been associated with more advanced stages in implementing AI governance actions at a strategic level.

The positive impact of the four studied practices aligns with the researchers' arguments that a risk management approach for AI requires that traditional risk management processes (Chen & Deng, 2022; Duijm, 2015) be associated with stakeholders' identification (Schaefer et al., 2021; Wirtz et al., 2022; Wright & Schultz, 2018), with a risk-oriented audit process (De Oliveira, 2019; Erlina et al., 2020), and with a change management in the environment (González et al., 2020).

4.1.3. Analysis of proposition 1C

Considering only the cyber security management process, this analysis did not present combinations of processes or practices because this process was not decomposed into other practices. The fuzzy QCA analysis (Table 4) revealed that, of the cases that had implemented a process for managing security in AI systems (PSEC), 66.67 % (cases 3, 5, 6, 8, 9, 13, 19, 22, 25, 26) obtained a high score for AI governance actions at a strategic level (SAIGOV(1)) (Appendix A - Annex 6), which indicates an association between the existence of a security management process and advanced stages of AI governance actions at a strategic level. Such result confirms researchers' arguments that a) AI governance implies, among other principles, ensuring robust and safe AI systems (Dalrymple et al., 2024), b) since robust and safe AI systems require practices to deal with mechanisms to face cyberattacks created specifically to AI systems (Booth et al., 2023; Ee et al., 2024), AI governance also requires practices systematically organized to manage the cyber security of AI systems. Compared with the other propositions' analyses, this was the lower consistency (0.8312) in the considered associations, which can be further investigated in future research.

4.1.4. Analysis of proposition 1D

The analysis of system development practices required two levels of decomposition: firstly, the phases of the AI system development and support process – project (Proposition 1D1) and operation (1D2); and subsequently, decomposing each of those phases.

4.1.4.1. Analysis of Proposition 1D1. The fuzzy QCA applied to the project phase practices (Table 5) used the representation of rules and ethical dilemmas (RDILEM), practices to minimize biases (PBIAS), and practices to provide transparency (TRANSP) in the AI system development (Appendix A - Annex 6). The low average score for practices aimed at transparency (50.11, see Appendix A - Annex 4) confirms the challenge of obtaining an explanation for the algorithm's results (Buiten, 2019; Butterworth, 2018; Zuiderwijk et al., 2021). Among other factors, this scenario may have been amplified by the fact that 73.33 % of the sample outsourced at least part of their AI system development, and only 46.43 % of the sample had access to their AI system code. The low average score for the representation of principles and ethical dilemmas (51.11, see Appendix A - Annex 4) may be due to the lack of a clear definition of those principles or the lack of specialists to implement the practice, as Ahn and Chen (2022) have alerted. Among the cases that had deployed practices to represent the business rules, principles and ethical dilemmas, and practices for transparency in AI system development, 75 % (cases 3, 9, 13, 21, 26, 27) showed high scores for AI governance actions at a strategic level (SAIGOV(1)).

When going deeper into the fuzzy-value analysis, we notice that those cases also implemented practices to minimize biases. The "necessary condition" test showed greater consistency (Cs = 0.96263) for the combination of the three practices. Therefore, the analysis of Proposition 1D1 revealed that in the studied sample, organizations that implemented practices to represent rules related to principles and ethical dilemmas, practices to provide transparency, and practices to minimize biases in AI systems during their development are associated with a more advanced stage in the implementation of AI governance actions at a strategic level.

Such result aligns with the researchers' arguments that, to ensure trustworthy AI systems, it is necessary to implement practices for

Table 3
Fuzzy QCA applied to Proposition 1B.

Proposition	Combinations	Cases	Results	Coverage and Consistency
1B	1. ~PAUDIT * STAKEH * ENVIRO	27,25	SAIGOV(1)	Cv:0.319031
		2,16	SAIGOV(0)	Cs:0.893411
	2. PRISM * STAKEH * ENVIRO	3,9,13,22,27,8	SAIGOV(1)	Cv: 0.49135
		2	SAIGOV(0)	Cs: 0.941645
	3.~PRISM*PAUDIT*STAKEH* ~ ENVIRO	5,21	SAIGOV(1)	Cv: 0.293426
				Cs: 0.925764

Table 4
Fuzzy QCA applied to Proposition 1C.

Proposition	Combinations	Cases	Results	Coverage and Consistency
1C	PSEC	5,8,9,13,22,3,6,19,25,26 11,2,16,20,28	SAIGOV(1) SAIGOV(0)	Cv:0.719031 Cs: 0.8312

Table 5
Fuzzy QCA applied to Proposition 1D1.

Proposition	Combinations		Cases	Results	Coverage and Consistency
1D1	RDILEM * TRANSP		9,13,27,3,21,26 15,24	SAIGOV(1) SAIGOV(0)	Cv: 0.538408 Cs: 0.913146
	"Necessary Conditions" Test	RDILEM * TRANSP		SAIGOV(1)	Cv: 0.727125 Cs: 0.929412
		RDILEM * TRANSP * PBIAS		SAIGOV(1)	Cv: 0.679531 Cs: 0.96263

Table 6
Fuzzy QCA applied to Proposition 1D2.

Proposition	Combinations		Cases	Results	Coverage and Consistency
1D2	HOVER * FEEDB		3,6,9,13,22,25,5,14, 19,26,27 10,24,16,20,28	SAIGOV(1) SAIGOV(0)	Cv: 0.741869 Cs: 0.790561
	"Necessary Conditions" Test	HOVER* FEEDB			Cv: 0.627069 Cs: 0.891349
		AMONI*HOVER*FEEDB			Cv: 0.576529 Cs: 0.972318

improving transparency along with the AI system development (Adadi & Berrada, 2018; Arrieta et al., 2020; Das, 2020; Dazeley et al., 2021; Kale et al., 2022; Phillips et al., 2021; Schaefer et al., 2021), for identification of ethical principles and dilemmas to be applied in the business rules (González et al., 2020; Rajkomar et al., 2018), and for mitigating biases in data preparation and in modeling (Ashokan & Haas, 2021; Baeza-Yates, 2018; De Silva & Alahakoon, 2022; Leavy et al., 2020; Lin et al., 2021; Makhlouf et al., 2021; Ntoutsi et al., 2020; Oneto & Chiappa, 2020; Silberg & Manyika, 2019).

4.1.4.2. Analysis of proposition 1D2. The analysis of practices in the operation phase of the AI system development and maintenance process (Table 6) considered the automatic monitoring (AMONI), human oversight (HOVER), and user feedback (FEEDB) variables (Appendix A - Annex 6). The highest average score was identified among the practices of this phase in automatic monitoring (71.57, see Appendix A - Annex 4), revealing the greater ease of monitoring when one does not depend on human resources. The preliminary fuzzy QCA showed a combination composed of human oversight and user feedback in sixteen organizations, of which 68.75 % (cases 3, 5, 6, 9, 13, 14, 19, 22, 25, 26, 27) obtained high scores for the AI governance actions at a strategic level (SAIGOV(1))). It is important to note that fifteen cases implemented automatic monitoring, human oversight, and user feedback, confirming the perceptions of Rahwan et al. (2019), Wright and Schultz (2018), and De Silva and Alahakoon (2022). Additionally, the "Necessary Conditions" test indicates that a high score would unlikely be obtained for AI governance actions at a strategic level without implementing the three practices (consistency = 0.972318). Therefore, there is an association between organizations at a more advanced stage in AI governance actions at a strategic level and cases that had automatic monitoring, human oversight, and user feedback practices, as argued by De Silva and Alahakoon (2022), Laato et al. (2022), Strauß (2021), González et al. (2020), and Zicari et al. (2021), when they demanded monitoring AI in the real environment.

4.2. Analysis of proposition 2

The analysis of Proposition 2 (Table 7) included the fuzzy QCA variables (Appendix A - Annex 6) training in data, AI risks, and AI ethical principles directed at decision-makers (TDEMAK); training in data, AI system development, AI risks, and AI ethical principles targeting developers (TDEVEL); training in data for users (TUSER); and training in data, AI risks and AI ethical principles for auditors (TAUDIT). Focusing on figuring out whether the mentioned trainings are enablers of AI governance practices, the overall score of actions towards AI governance was considered as the outcome variable (actions at the strategic level, ancillary processes and practices, and practices that belong to the AI system development process) (GAIGOV(1)).

Delivering a consistency of 0.937173, the fuzzy QCA revealed that the combination characterized by training aimed at decision-makers, developers, and users showed 87.5 % (cases 8, 9, 13, 22, 24, 27, 28) of these cases with high scores for actions focused on AI governance. Such result indicates an association between training key stakeholders and higher stage implementation of AI governance, as argued by Calzada & Almirall, 2020; Micheli et al., 2020; Ruijer, 2021; Makarius et al., 2020; Herremans, 2021, Pinski et al., 2024; Schüller, 2022.

Table 7
Fuzzy QCA applied to Proposition 2.

Proposition	Combinations	Cases	Results	Coverage and Consistency
2	1. TDEVEL * ~ TAUDIT	27,19,20,28 14,25,26	GAIGOV (1) GAIGOV (0)	Cv: 0.452348 Cs: 0.927762
	2. TDEMAK * TDEVEL * TUSER	13,22,27,8,9,24,28	GAIGOV (1)	Cv: 0.494475 Cs: 0.937173

Table 8
Crisp-set Fuzzy QCA applied to Proposition 3.

Proposition	Combinations	Cases	Results
3	1. TUSEAI{1}*DTDEMAK{0}* DTDEVEL{0}	6,3	HAIGOV(1)
		1,4,2,10,15,18,21,23	HAIGOV(0)
	2. ACOD80{0}*TUSEAI{0}* DTDEMAK{1}	5,20,28	HAIGOV(1)
		22	HAIGOV(0)
	3. ACOD80{0}*TUSEAI{0}* DTDEVEL{1}	19,20,28	HAIGOV(1)
		22	HAIGOV(0)
	4. TUSEAI{1}*DTDEMAK{1}* DTDEVEL{1}	8,9,13,27,24,25	HAIGOV(1)
		26	HAIGOV(0)

The absence of auditor training revealed that focus on internal audits for AI systems has not been a priority for those organizations, although a small group has provided the four trainings.

4.3. Analysis of proposition 3

Proposition 3 was analyzed through a crisp-set QCA (Table 8) using the following dichotomous variables (Appendix A - Annex 6): training for decision-makers (DTDEMAK), training for developers (DTDEVEL), the organization's access to at least 80 % of their AI system code (ACOD80) (four of the five AI systems reported in the questionnaire), and more than three years of experience in AI system development (TUSEAI). For the outcome variable, a dichotomous variable was created to indicate the sum of the fuzzy scores of all practices for implementing AI governance (HAIGOV). HAIGOV = 1, if = GAIGOV≥60, and HAIGOV = 0 if GAIGOV<60.

Combination 1 — consisting of cases with more than three years in AI system production, without training for decision-makers and for AI system developers — was associated with low (not high) overall scores for all actions towards AI governance (cases 1, 2, 4, 10, 15, 18, 21, 23). At the other end, Combination 4 — comprising cases with more than three years of AI system production that have offered training to decision-makers and AI system developers — was associated with more advanced stages of the implementation of AI governance (cases 8, 9, 13, 24, 25, 27). Combinations 2 and 3 were not assertive enough to any conclusion.

It is worth observing that having (or not) access to the AI system codes, which implies outsourcing AI systems, did not impact Combinations 1 and 4. According to Combination 4, training decision-makers and developers can be associated with an advanced stage in implementing AI governance practices. And according to Combination 1, not training decision-makers or developers can be associated with lower stages in implementing AI governance. Both situations align with Ahn and Chen (2022) and Benfeldt et al. (2020). Preposition 3 reinforces the need to train key stakeholders even when the public organization outsources the development of AI systems.

5. Analysis of the interview responses and documents

The analysis of the interview responses and documents provided by the organizations was carried out while attempting to understand how practices and processes categorized in the research model were applied and used in the analysis of propositions.

5.1. Processes and practices for AI governance

In the context of actions, processes, and practices, many approaches were given and challenges were found in the path towards AI governance in the studied public organizations (Appendix A - Annex 7)(CNJ, 2020; LIAA-3R, 2022; Nagbøl & Müller, 2020; Nagbøl, Müller, & Krancher, 2021; Vero, 2019). Since all efforts towards AI governance are distributed at many levels of the organization hierarchy, it reflects the organization's culture and its risk appetite.

5.2. Government standards and guidelines

Along with the interviews, one observes that organizations whose governments have already established AI ethical principles guidelines for all their agencies have adopted those principles completely and, in a few cases, they have added details to their policies or strategies to customize the guidelines to their singularities (See "Policies and Guidelines for AI Ethical Principles" in Appendix A - Annex 7 and Appendix A - Annex 9). Similarly, in some cases, governments create agencies specialized in developing standards for processes and practices related to AI governance (See Appendix A - Annex 8). In both cases, those organizations saved time, money, and human resources, as one can observe in Appendix A - Annex 10.

Regarding the construction of standards and transfer of knowledge, it is worth highlighting a partnership between the public and private sectors established by the Finnish Government (Aurora, 2019) and an agreement between Nordic countries to implement best practices for AI systems with a focus on ethical issues (Nordic Council of Ministers, 2018).

5.3. Outsourcing

In 73.33 % of the organizations, third parties were hired or partnered with to develop at least part of their AI systems. Considering that public organizations are not self-sufficient to produce AI systems on the scale and with the level of expertise they require, according to Hickok (2022), Zick et al. (2024) and Coglianese (2024), outsourcing is also a driver to implement practices for AI governance (Appendix A - Annex 10). A few organizations were inspired by the World Economic Forum's model (WEF, 2020, 2020b) for outsourcing AI system development compliant with ethical principles.

5.4. Partnership between the business unit and the IT unit

Among the interviewees, there is the perception of the "business + IT" joint action as a strategy to minimize biases, provide transparency, and implement data governance. A second group of reports was provided by professionals who implemented the risk management process and the AI system development process, with artifacts filled out by the IT and business staff (Nagbøl et al., 2021). In Annex 10, one can find some of the interviewees' responses and comments regarding the "business + IT" partnership.

6. Merging the analyses

6.1. Associations found in the QCA

The proposition results are summarized as follows:

Proposition 1 results: An association was found between AI governance actions at the strategic level ("create an AI governance structure," "elaborate an AI strategy," "establish an AI policy," "implement an AI governance process," "establish an AI ethical principles code") and the existence of data-related processes (data governance, data quality management, personal data protection management), risk-related processes and practices (risk management process, stakeholder definitions, monitoring changes in the environment), security management processes, AI system development practices (rules representing ethical principles and ethical dilemmas, biases minimization, transparency, automatic monitoring, human oversight, and feedback collection). No association was found between the audit process and the strategic actions for AI governance.

Proposition 2 results: An association was found between advanced stages of AI governance implementation and training targeting decision-makers, AI system developers, and digital services users, confirming that training such stakeholders is a driver for implementing AI governance.

Proposition 3 results: In the context of organizations with more than three years of experience in AI system production, regardless of the make-or-buy decision, training decision-makers and developers is associated with an advanced stage of implementing all AI governance practices. And not training decision-makers or developers is associated with lower stages in implementing AI governance. Thus, regardless of the make-or-buy decision, training is a driver of AI governance implementation.

6.2. Relevant contributions from the government to the whole public sector

Interviews and documents made available by governments showed the benefits of having a central government body that produces clear and accessible guidelines with recommendations for AI system development, which confirms Mikalef et al.'s (2022b) and Schaefer et al.'s (2021) perceptions. Some guides made up the governments' portfolio of standards for promoting AI governance in agencies and departments under their responsibility (Australian Government, 2019a, 2019b; Norwegian Data Protection Authority, 2018; Government of United Kingdom, 2017, 2019b, 2020a, 2020b, 2020c, 2020d, 2020e, 2021a, 2021b, 2021c, 2021d, 2021e, 2022a, 2022b, 2022c, 2022d, 2022e; Information Commissioner's Office, 2020, 2021; Ekspertgruppen om dataetik, 2018; Balahur et al., 2022; German Federal Ministry for Economic Affairs and Energy, 2020; AI HLEG, 2019; Government of Canada, 2019a, 2019b, 2020a, 2020b, 2021a, 2021b, 2022a, 2022b; Leslie, 2019; National Institute of Advanced Industrial Science and Technology, 2022; European Union Agency for Cyber Security, 2021; WEF, 2020a, 2020b; C4IR Brasil, 2022; Switzerland Federal Council, 2021, Council of Europe, 2021, European Data Protection Board, 2022) (Appendix A - Annex 8). In addition to supplying knowledge that is lacking in many public organizations, these specialized institutions speed up implementation and promote a standard for concepts that facilitates communication among government departments. In a strategic context, some governments (German Federal Government, 2020;, Presidencia de

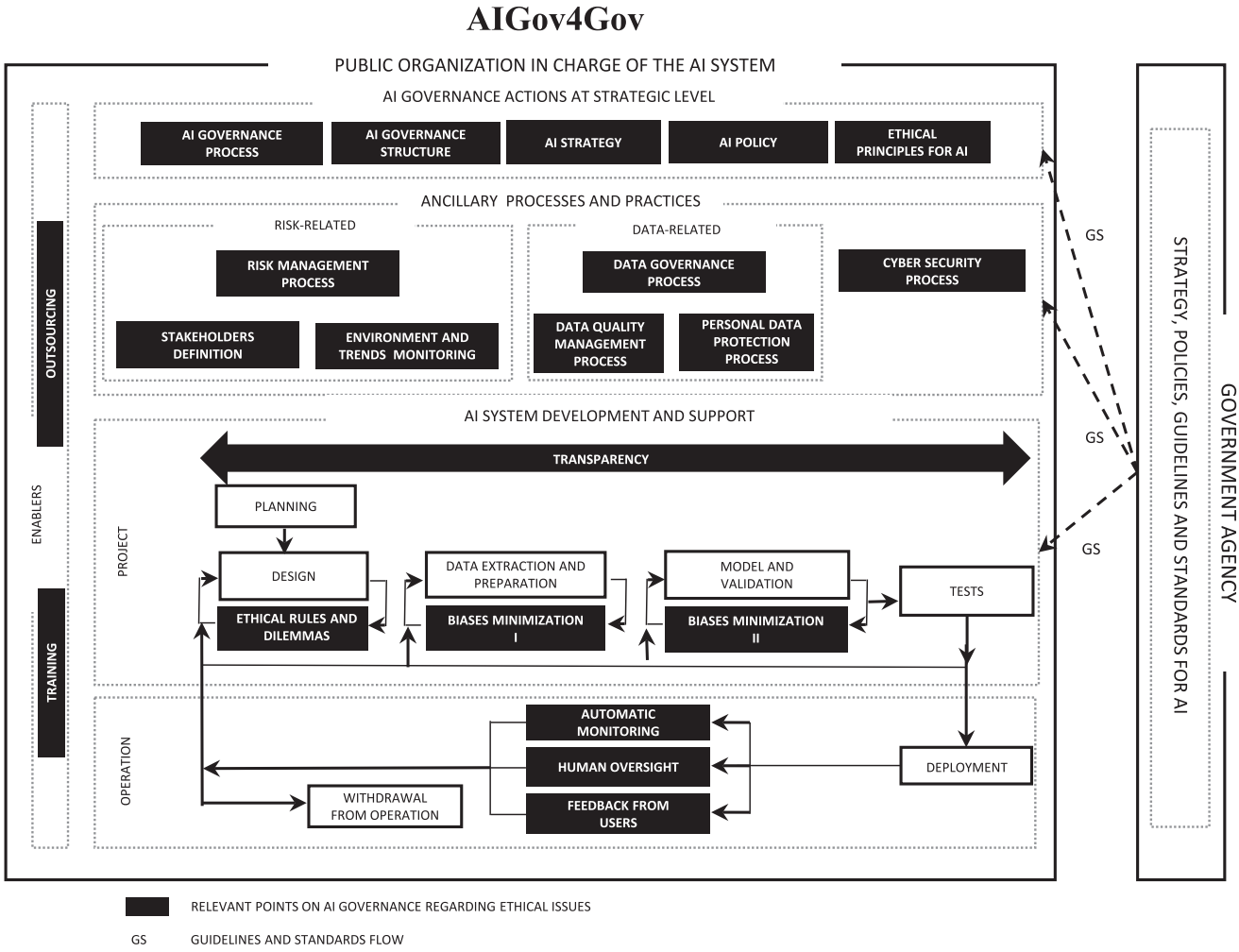


Fig. 6. AIGov4Gov – AI governance framework proposed for public organizations. (Source: Self-elaboration)

la Nación, 2019; Australian Government, 2021; Ministério da Ciência Tecnologia e Inovação do Brasil, 2021; Government of Canada, 2022b; Danish Government, 2019; Government of the Republic of Estonia, 2019; Ministry of Economic Affairs and Employment of Finland, 2017, 2019; Ministero dello sviluppo economico, 2019; Japanese Strategic Council for AI Technology, 2017; Ekspertgruppen om dataetik, 2018; Government of the Grand Duchy of Luxembourg, 2018; Norwegian Ministry of Local Government and Modernisation, 2020; Government of United Kingdom, 2021e; Government of Sweden, 2020; European Commission, 2018a) assign AI strategic guidelines to an agency that develops and monitors the national AI strategy and establishes policy and ethical principles for AI systems (Appendix A - Annex 9). In both cases, public organizations can not only adopt the government's central strategic guidelines and standards but can also build their own versions in line with the government's general definitions.

6.3. Drivers to AI governance implementation

Combining the findings of Propositions 2 and 3 with the interviews, training key stakeholders and outsourcing were considered drivers of AI governance implementation in public organizations.

6.4. AIGov4Gov framework

Consolidating the QCA results and the interviews' results, a conceptual view was built in a multilevel approach encompassing practices and processes found in the sample aimed at the production of AI systems that considered ethical principles, drivers to implement AI governance, and the Government Central Agency contributions to the public sector: the AIGov4Gov framework (Fig. 6).

Reaching the strategic, tactical, and operational levels, AIGov4Gov is an AI governance model for public organizations. At the strategic level, there are actions aimed at elaborating an AI strategy, establishing an AI policy, establishing an AI ethical principles code, implementing an AI governance process, and creating an AI governance structure for said governance. To support those actions, ancillary processes and practices are combined for a) data governance, supported by data quality management and personal data protection management; b) AI-related risk mitigating using a risk management process supported by a stakeholders' definition and by an environment and social trends monitoring in line with an audit process; c) Cyber security management. The expected increase of legislation to regulate AI worldwide was considered when deciding to maintain the audit process within the framework. At the tactical and operational levels, AIGov4Gov provides practices for AI system development and maintenance processes with a focus on ethical principles and aligned to researchers' argument for an AI system development process based on agile methods, which use continuous loops in each phase (Laato et al., 2022), and which consider ethical principles in the loops (Leijnen et al., 2020; Lu et al., 2024). Right after planning, development takes place in several interactions during problem specification and with the representation of ethical principles and dilemmas. Then, practices for minimizing biases are applied in successive iterations during the data extraction and preparation phase, as well as during the model construction and validation stage, followed by testing. Soon after deployment, practices to follow-up AI systems are implemented in the complexity of the real environment through automatic monitoring, human oversight, and collection of user feedback.

As an enabler of AI governance implementation, AIGov4Gov provides outsourcing in addition to training for key stakeholders, like decision-makers, developers, and users, customized for their role in the implemented processes and practices. Auditor training can also be considered when aiming for a scenario where AI legislation is a comprehensive reality.

Also included in AIGov4Gov is the interaction between the public organization in charge of the AI system and the agency in its sphere of government (if any) in charge of AI strategy, AI policy, and AI ethical principles for AI systems applicable to organizations under its responsibility. In such a situation, the organization in charge of the AI system can adopt its centralized government strategic guidelines or adapt them while complying with them. Similarly, guidelines are provided with standards for risk management, data governance, data quality management, personal data protection management, cyber security management, minimization of biases, and transparency throughout the AI system development process (GS Flow in Fig. 6). When the organization decides to outsource, the AI system development process still requires the "business + IT" partnership to enable the ancillary processes and practices for AI governance.

7. Conclusions

This study investigated how public organizations have incorporated the guidelines presented by academia, international standards, and legislation for AI system development, considering ethical principles in their governance and management processes. The results confirmed the perception that AI governance requires a multilayer model with strategic-level actions that guide processes and ancillary AI governance practices in a combined action. All processes and practices designed in the research model were observed in the sample. However, data-related, risk-related, and AI system development processes and practices were prioritized in the sample for AI governance implementation. The cyber security management process received lower adherence, and audit processes were still seldom adopted at the time of data gathering when very few countries had approved laws to regulate AI. Regarding development in the project phase, organizations at a more advanced stage in AI governance have prioritized practices representing ethical principles and ethical dilemmas, transparency practices, or practices to minimize biases. When those AI systems are in operation, organizations at a more advanced stage in AI governance have implemented practices for automatic monitoring, human oversight, and collection of user feedback.

Training key stakeholders and outsourcing are enablers of an AI governance implementation. It was observed that organizations that have outsourced their AI system development have also trained managers as well as AI developers, and the lack of training of those professionals is associated with less advanced stages in AI governance.

One could observe the huge opportunity that government agencies have to promote AI governance by defining guidelines for all organizations under their responsibility or recommending standards for AI governance despite the long time required to deploy them. In the context of public organizations, it is worth mentioning the need for policy-making that combines an internal multilayered approach with a continuous alignment with the government guidelines and standards. The findings also corroborated to a proposed framework — AIGov4Gov — encompassing combined processes and practices to establish AI governance in a public organization.

8. Limitations and agenda

Limitations found in this research:

a) Despite the broad and systematic process of obtaining the sample, this research carried out analyses in countries and organizations that published their AI systems, made contacts available, and agreed to participate in the research. Therefore, despite efforts to include representatives from all countries with a high level of AI system production, the sample does not follow the global AI ranking proportions, nor does it have a balanced representation of each continent.

b) The research focused on capturing the existence of practices and processes but did not delve into each process and its maturity model. Thus, each participant had his own perception regarding whether a practice/process was being implemented completely or partially. For the same reason, the research did not make a deeper analysis of the quality of the training offered to users, developers, decision-makers, and auditors.

c) Similar to the previous item, the government AI standards found in the sample were not classified considering their maturity.

d) The research encompassed only practices and processes representing all efforts to implement AI governance. Impacts on corporate governance and e-government were out of the scope.

e) The AI systems presented by the organizations in the sample did not consider generative AI, probably because the gathering criteria were systems that were in operation and were already part of the organization's official portfolio when the data were collected.

As an exploratory and descriptive study, this research paves the way for an agenda of new investigations that go deeper into the findings regarding each proposition, as well as an investigation of how practices and processes studied under the AI governance effort would impact corporate governance and e-government. Finally, a space is opened for deepening the AI system lifecycle through a maturity model for AI system development and support focusing on ethical principles.

9. Contribution

This research presents itself as innovative in terms of content, as it addressed the gap highlighted by Mäntymäki et al. (2022) and Mikalef, Conboy, et al. (2022) in the empirical knowledge of how organizations have interpreted and incorporated AI system development best practices into their processes and practices. The research has innovated by using crisp-set QCA, fuzzy QCA, and content analyses as it addressed the gaps presented by Zuiderwijk et al. (2021). Therefore, the following contributions to managers and researchers can be summarized: a) identification of how processes and practices aimed at applying ethical principles in AI system development have been combined and internalized in the governance and management models of public organizations; b) identification of how AI governance enablers have been used by public organizations; c) how a central government agency can booster AI governance in government agencies; d) a framework for AI governance in public organizations, in which processes and practices are articulated at the strategic, tactical, and operational levels in AI system production that consider ethical principles.

CRediT authorship contribution statement

Patricia Gomes Rêgo de Almeida: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Carlos Denner dos Santos Júnior:** Supervision, Writing- original draft, Investigation, Visualization, Validation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.giq.2024.102003.

References

Aasi, P., Rusu, L., & Han, S. (2014). The influence of culture on IT governance: A literature review. In *47th Hawaii international conference on system sciences* (pp. 4436–4445). https://doi.org/10.1109/HICSS.2014.546

Abdollahi, B., & Nasraoui, O. (2018). Transparency in Fair machine learning: The case of explainable recommender systems. In J. Zhou, & F. Chen (Eds.), *Human and Machine Learning. Human–Computer Interaction Series*. Cham: Springer. https://doi.org/10.1007/978-3-319-90403-0_2.

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management, 49*(July), 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

Aburachid, L. M. C., & Greco, P. J. (2011). Validação de conteúdo de cenas do teste de conhecimento tático no tênis. *Estudos de Psicologia. Campinas, 28*(2), 261–267.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access Review, 6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Agarwal, L.. Defining organizational AI governance and ethics. Available at SSRN: https://ssrn.com/abstract=4553185.

Ahn, M. J., & Chen, Y. (2022). Digital transformation toward AI-augmented public administration: The perception of government employees and the willingness to use AI in government. *Government Information Quarterly, 39), Issue 2*. https://doi.org/10.1016/j.giq.2021.101664

AI HLEG. (2019). Ethics guidelines for trustworthy AI. European Commission. Retrieved from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

Aiken, C. (2021). *Classifying AI systems. Center for Security and Emerging Technology*. Georgetown University. Retrieved from https://cset.georgetown.edu/publication/classifying-ai-systems/ Accessed September 23, 2024.

Alhosani, K., & Alhashmi, S. M. (2024). Opportunities, challenges, and benefits of AI innovation in government services: A review. *Discover Artificial Intelligence, 4*, 18. https://doi.org/10.1007/s44163-024-00111-w

Alshahrani, A., Dennehy, D., & Mäntymäki, M. (2021). An attention-based view of AI assimilation in public sector organizations: The case of Saudi Arabia. *Government Information Quarterly, 39), Issue 4*. https://doi.org/10.1016/j.giq.2021.101617

Anderson, M., & Anderson, S. L. (2018). Geneth: A general ethical dilemma analyzer. De Gruiter. Paladyn. *J. Behav. Robot., 9*, 337–357. https://doi.org/10.1515/pjbr-2018-0024

Andrews, L. (2018). Public administration, public leadership and the construction of public value in the age of the algorithm and 'big data'. *Public Administration, 97*(2), 296–310. https://doi.org/10.1111/padm.12534

Arrieta, A. B., Díaz-Rodríguez, N., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., … Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for artificial intelligence and digital technologies. *International Journal of Information Management, 62*. https://doi.org/10.1016/j.ijinfomgt.2021.102433

Ashokan, A., & Haas, C. (2021). Fairness metrics and bias mitigation strategies for rating predictions. *Information Processing & Management, 58*(5), Article 102646. ISSN 0306–4573 https://doi.org/10.1016/j.ipm.2021.102646.

Aurora, A. I. (2019). Aurora AI - Towards a human Centric Society. Retrieved from https://vm.fi/documents/10623/1464506/AuroraAI+development+and+implementation+plan+2019%E2%80%932023.pdf.

Australian Government. (2019). *Artificial intelligence – Australia's ethics framework*. Innovation and Science: Department of Industry. Retrieved from https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework Accessed September 23, 2024.

Australian Government. (2019b). Australian Privacy Principles Guidelines. Office of Australian Information Commissioner. Retrieved from https://www.oaic.gov.au/_data/assets/pdf_file/0009/1125/app-guidelines-july-2019.pdf Accessed September 23, 2024.

Australian Government. (2021). Australian's Artificial Intelligence Action Plan. Retrieved from https://www.industry.gov.au/publications/australias-artificial-intelligence-action-plan.

Awad, E., Dsouza, S., Bonnefon, J. F., Shariff, A., & Rahwan, I. (2020). Crowdsourcing moral machines. *Communications of the ACM, 63*(3), 48–55. https://doi.org/10.1145/3339904

Baeza-Yates, R. (2018). Bias on the web. *Communications of the ACM, 61*(6), 54–61. https://doi.org/10.1145/3209581

Balahur, A., Jenet, A., Torres, I., Charisi, V., Ganesh, A., Griesinger, C. B., Maurer, P., Mian, L., Salvi, M., Scalzo, S., Sol er Garrido, J., Taucer, F., & Tolan, S. (2022). *Data quality requirements for inclusive, non-biased and trustworthy AI. Putting-Science-Into-Standards*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2760/365479

Bench-Capon, T., & Modgil, S. (2017). Norms and value based reasoning: Justifying compliance and violation. *Artificial Intelligence & Law Review, 25*, 29–64. https://doi.org/10.1007/s10506-017-9194-9

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data governance as a collective action problem. *Information Systems Frontiers, 22*, 299–313. https://doi.org/10.1007/s10796-019-09923-z

Betarelli-Júnior, A. A., & Ferreira, S. F. (2018). *Introdução à Análise Qualitativa Comparativa e aos Conjuntos Fuzzi (FSQCA)*. Enap: Brasilia.

Bonnemains, V., Saurel, C., & Tessier, C. (2018). Embedded ethics: Some technical and ethical challenges. *Ethics and Information Technology, 20*, 41. https://doi.org/10.1007/s10676-018-9444-x

Bonsón, E., Lavorato, D., Lamboglia, R., & Mancini, D. (2021). Artificial intelligence activities and ethical approaches in leading listed companies in the European Union. *International Journal of Accounting Information Systems, 43*. https://doi.org/10.1016/j.accinf.2021.100535, 1467–0895.

Booth, J., Metz, D. W., Tarkhanyan, D. A., & Cheruvu, S. (2023). Machine learning security and trustworthiness. In *Demystifying intelligent multimode security systems*. Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4842-8297-7_5.

Boyd, R., & Holton, R. J. (2018). Technology, innovation, employment and power: Does robotics and artificial intelligence really mean social transformation? *Journal of Sociology, 54*(3), 331–345.

Breier, J., Baldwain, A., Balinsky, & Liu, Y. (2020). Risk Management for Machine Learning Security. *arXiv*. https://doi.org/10.48550/arXiv.2012.04884, 2012.04884v1 [cs.CR].

Buiten, C. M. (2019). Towards intelligent regulation of artificial intelligence. *European Journal of Risk Regulation, 10*(1), 41–59. https://doi.org/10.1017/err.2019.8

Butterworth, M. (2018). The ICO and artificial intelligence: The role of fairness in the GDPR framework. *Computer Law and Security Review, 34*, 257–268. https://doi.org/10.1016/j.clsr.2018.01.004

C4IR Brasil. (2022). Guia de Contratações Públicas de Inteligência Artificial. Centro para a 4ª Revolução Industrial. Retrieved from https://c4ir.org.br/wp-content/uploads/2022/11/1648128585465GUIA-DE-CONTRATACOES-PUBLICAS-DE-AI_C4IR_v4.pdf Accessed September 23, 2024.

Calzada, I., & Almirall, E. (2020). Data ecosystems for protecting European citizens' digital rights. *Transforming Government: People, Process and Policy, 14*(2), 133–147. https://doi.org/10.1108/TG-03-2020-0047

Carretero, A., Gualo, F., Caballero, I., & Piattini, M. (2017). *MAMD 2.0: Environment for data quality processes implantation based on ISO 8000-6X and ISO/IEC 33000. 54* pp. 139–151). Elsevier BV. https://doi.org/10.1016/j.csi.2016.11.008

Carter, D. (2020). (2020). Regulation and ethics in artificial intelligence and machine learning technologies: Where are we now? Who is responsible? Can the information professional play a role? *Business Information Review, 37*(2), 60–68. https://doi.org/10.1177/0266382120923962

Cerka, P., Grigiene, J., & Sirbikyte, G. (2017). Is it possible to grant legal personality to artificial intelligence software systems? *Computer Law and Security Review, 33*(5), 685–699.

Chen, T., Liu, J., Xiang, Y., Niu, W., Tong, E., & Han, Z. (2019). Adversarial attack and defense in reinforcement learning-from AI security view. *Cybersecurity, 2*, 1–22. https://doi.org/10.1186/s42400-019-0027-x

Chen, X., & Deng, Y. (2022). An evidential software risk evaluation model. *Mathematics, 10*, 2325. https://doi.org/10.3390/math10132325

Cihon, P., Maas, M. M., & Kempo, L. (2020). Should artificial intelligence governance be centralized?: Design lessons from history. In *AAAI/ACM conference on AI, ethics, and society (AIES '20)* (pp. 228–234). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3375627.3375857.

CNJ. (2020). Resolução N° 332 de 21/08/2020. Retrieved from https://atos.cnj.jus.br/atos/detalhar/3429 Accessed September 23, 2024.

Codá, R. C., Farias, J. S., & Dias, C. (2022). Interactive value formation and lessons learned from Covid-19: The Brazilian case. *Journal of Quality Assurance in Hospitality and Tourism*. https://doi.org/10.1080/1528008X.2022.2135057

Coglianese, C. (2024). Procurement and artificial intelligence. In *Handbook on public policy and artificial intelligence* (pp. 235–248). Edward Elgar Publishing. https://doi.org/10.4337/9781803922171.00026.

Council of Europe. (2021). Guidelines on facial recognition. Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data. Convention 108. Retrieved from https://rm.coe.int/guidelines-facial-recognition-web-a5-2750-3427-6868-1/1680a31751 Accessed September 23, 2024.

Dalrymple, D., Skalse, J., Bengio, Y., Russell, S., Tegmark, M., Seshia, S., Omohundro, S., Szegedy, C., Goldhaber, B., Ammann, N., Abate, A., Halpern, J., Barrett, C., Zhao, D., Zhi Xuan, T., Wing, J., & Tenenbaum, J. (2024). *Towards Guaranteed Safe AI: A Framework for Ensuring Robust and Reliable AI Systems*. arXiv:2405.06624v3 [cs.AI]. https://doi.org/10.48550/arXiv.2405.06624.

Danish Government. (2019). *National Strategy for Artificial Intelligence*. Ministry of Finance and Ministry of Industry, Business and Financial Affairs. Retrieved from https://en.digst.dk/strategy/the-danish-national-strategy-for-artificial-intelligence/ Accessed September 23, 2024.

Danish Government. (2020). Lov om ændring af årsregnskabsloven. (Krav om rapportering af dataetik). Retrieved from https://www.retsinformation.dk/eli/lta/2020/741 Accessed September 23, 2024.

Das, A. (2020). *Opportunities and challenges in explainable artificial intelligence 9XAI: A survey*. arXiv:2006.11371 [cs.CV]. https://doi.org/10.48550/arXiv.2006.11371.

Dazeley, R., Vamplew, P., Foale, C., Young, C., Aryal, S., & Cruz, F. (2021). Levels of explainable artificial intelligence for human-aligned conversational explanations. *Artificial Intelligence, 299, 103525*. https://doi.org/10.1016/j.artint.2021.103525. ISSN 0004-3702.

De Almeida, P. G. R., dos Santos, C. D., & Farias, J. S. (2021). Artificial intelligence regulation: A framework for governance. *Ethics and Information Technology, 23*, 505–525. https://doi.org/10.1007/s10676-021-09593-z

De Oliveira, T. F. (2019). *Avaliação das Práticas de Auditoria Interna da Secretaria Federal de Controle Interno da CGU sob a Ótica da Auditoria Baseada em Riscos*. Brasil: Controladoria Geral da União. Retrieved from https://revista.cgu.gov.br/Revista_da_CGU/article/view/73/pdf_60 Accessed September 23, 2024.

De Silva, D., & Alahakoon, D. (2022). An artificial intelligence life cycle: From conception to production. *Patterns, 3*(6), Article 100489. https://doi.org/10.1016/j.patter.2022.100489

Dennis, L., Fisher, M., Slavkovik, M., & Webster, M. (2016). Formal verification of ethical choices in autonomous systems. *Robotics and Autonomous Systems, 77*, 1–14. ISSN 0921-8890, https://doi.org/10.1016/j.robot.2015.11.012.

Dias, O. C. (2011). *Análise Qualitativa Comparativa (QCA) Usando Conjuntos Fuzzy – Uma Abordagem Inovadora Para Estudos Organizacionais no Brasil*. Rio de Janeiro: XXXV Encontro da ANPAD.

Dignum, V. (2019). AI is multidisciplinar. *AI Matters, 5*(4), 19–21. https://doi.org/10.1145/3375637.3375644

Djeffal, C. (2018). *Sustainable AI Development (SAID): On the Road to More Access to Justice*. https://doi.org/10.2139/ssrn.3298980

Duijm, N. J. (2015). Recommendations on the use and design of risk matrices. *Safety Science, 76*, 21–31. ISSN 0925-7535. https://doi.org/10.1016/j.ssci.2015.02.014

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., … Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57*, Article 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Ee, S., O'Brien, J., Williams, Z., El-Dakhakhni, A., Aird, M., & Lintz, A. (2024). Adapting cybersecurity frameworks to manage frontier AI risks: A defense-in-depth approach. *arXiv*. https://doi.org/10.48550/arXiv.2408.07933, 2408.07933v1 [cs.CY].

Eggers, S., & Sample, C. (2020). *Vulnerabilities in artificial intelligence and machine learning applications and data*. Idaho National Laboratory. US. Retrieved from https://inldigitallibrary.inl.gov/sites/sti/sti/Sort_57369.pdf Accessed September 23, 2024.

Eke, D. O., Chintu, S. S., & Wakunuma, K. (2023). Towards shaping the future of responsible AI in Africa. In D. O. Eke, K. Wakunuma, & S. Akintoye (Eds.), *Responsible AI in Africa. Social and cultural studies of robots and AI*. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-031-08215-3_8.

Ekspertgruppen om dataetik. (2018). Data i menneskets tjeneste Anbefalinger fra Ekspertgruppen om dataetik. Retrieved from https://www.em.dk/media/12013/ekspertgruppens-afrapportering-inkl-anbefalinger_final-a.pdf Accessed September 23, 2024.

Erlina, E., Nasution, A. A., Yahy, I., & Atmanegara, A. W. (2020). The role of risk based internal audit in improving audit quality. Erlina, Abdillah Arif Nasution, Idhar Yahya and Agung Wahyudhi Atmanegara, The Role of Risk Based Internal Audit in Improving Audit Quality. *International Journal of Management, 11*(12), 299–310.

European Commission. (2018). AI Watch. Retrieved from https://ai-watch.ec.europa.eu/index_en Accessed September 23, 2024.

European Commission. (2018a). Artificial Intelligence for Europe. Retrieved from https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence Accessed September 23, 2024.

European Commission. (2021a). Regulation of the European Parliament and of the Council Laying down harmonized rules on artificial intelligence. Artificial Intelligence Act and amending certain union legislative acts. Brussels. Retrieved from https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF Accessed September 23, 2024.

European Commission. (2021b). Selected AI cases in the public sector. Joint Research Centre [Dataset] PID. Retrieved from http://data.europa.eu/89h/7342ea15-fd4f-4184-9603-98bd87d8239a Accessed September 23, 2024.

European Data Protection Board. (2022). Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement. Version 1.0 Retrieved from https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-052022-use-facial-recognition_en Accessed September 23, 2024.

European Parliament. (2022a). *Draft Report on the proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM2021/0206 – C9–0146/2021–2021/0106(COD))*.

European Parliament. (2022b). Regulatory divergences in the draft AI act: Differences in public and private sector obligations, Study, European Parliamentary Research Service (EPRS), Brussels. Retrieved from https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729507/EPRS_STU(2022)729507_EN.pdf Accessed September 23, 2024.

European Union Agency for Cyber Security. (2021). *Securing Machine Learning Algorithms*. ISBN: 978–92–9204-543-2. https://doi.org/10.2824/874249.

European Union Agency for Cyber Security. (2022). Risk Management Standards – Analysis of standardisaton requeriments in supporting of cybersecurity policy. Retrieved from https://www.enisa.europa.eu/publications/risk-management-standards Accessed September 23, 2024.

FCT. (2021). Research in Data Science and Artificial Intelligence applied to Public Administration. Retrieved from https://www.fct.pt/wp-content/uploads/2022/06/Brochura_ResearchinDataScienceandAIappliedtoPA.pdf Accessed September 23, 2024.

Fernandes, G., Domingues, J., Tereso, A., & Pinto, E. (2021). A Stakeholders' perspective on risk Management for Collaborative. University-industry R&D programs. *Procedia Computer Science, 181*, 110–118. ISSN 1877-0509. https://doi.org/10.1016/j.procs.2021.01.110

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI* (p. 1). Berkman Klein Center Research Publication.

Freitas, V. S., & Neto, F. B. (2016). Qualitative Comparative Analysis (QCA): usos e aplicações do método. *Revista Política Hoje, 2ª Edição, 24*, 103–117.

Gerkensmeier, B., & Ratter, B. M. W. (2018). Governing coastal risks as a social process—Facilitating integrative risk management by enhanced multi-stakeholder collaboration. *Environmental Science & Policy, 80*, 144–151. ISSN 1462-9011. https://doi.org/10.1016/j.envsci.2017.11.011

German Federal Government. (2020). Artificial intelligence strategy of the German Federal Government. Retrieved from https://www.ki-strategie-deutschland.de/home.html Accessed September 23, 2024.

German Federal Ministry for Economic Affairs and Energy. (2020). German Standardization Roadmap on Artificial Intelligence. Retrieved from https://www.din.de/resource/blob/772610/e96c34dd6b12900ea75b460538805349/normungsroadmap-en-data.pdf Accessed September 23, 2024.

Gianni, R., Lehtinen, S., & Nieminen, M. (2022). Governance of responsible AI: From ethical guidelines to cooperative policies. *Frontiers in Computer Science, 4*, Article 873437. https://doi.org/10.3389/fcomp.2022.873437

González, F., Ortiz, T., & Ávalos, R. S. (2020). Responsible use of AI for public policy: Data science toolkit. In *OECD e IDB*. Retrieved from https://publications.iadb.or

g/publications/english/document/Responsible-use-of-AI-for-public-policy-Data-science-toolkit.pdf Accessed September 23, 2024.
Government of Canada. (2019a). Directive on Identity Management. Retrieved from https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=16577 Accessed September 23, 2024.
Government of Canada. (2019b). Directive on Security Management. Retrieved from https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32611 Accessed September 23, 2024.
Government of Canada. (2020). Guidelines on Service and Digital. Retrieved from https://www.canada.ca/en/government/system/digital-government/guideline-service-digital.html#ToC4_5 Accessed September 23, 2024.
Government of Canada. (2020a). Algorithmic Impact Assessment. Retrieved from https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html Accessed September 23, 2024.
Government of Canada. (2021). Directive on Automated Decision-making. Retrieved from https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592 Accessed September 23, 2024.
Government of Canada. (2021d). Levels of security. National Security and defense. Retrieved from https://www.tpsgc-pwgsc.gc.ca/esc-src/protection-safeguarding/niveaux-levels-eng.html.
Government of Canada. (2022a). Responsible use of artificial intelligence – Our guiding principles. Retrieved from https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html#toc1 Accessed September 23, 2024.
Government of Canada. (2022). Pan-Canadian Artificial Intelligence Strategy. Retrieved from https://ised-isde.canada.ca/site/ai-strategy/en Accessed September 23, 2024.
Government of India. (2020). Artificial Intelligence – Use Case Compendium. Ministry of Housing and Urban Affairs. Retrieved from https://iccc.smartcities.gov.in/pdf/AI-Use-Case-Compendium.pdf Accessed September 23, 2024.
Government of Sweden. (2020). AI Sweden. Retrieved from https://www.ai.se/en/news/swedish-government-establishes-new-ai-commission Accessed September 23, 2024.
Government of the Grand Duchy of Luxembourg. (2018). Artificial Intelligence: a strategic view for Luxembourg. Retrieved from https://innovative-initiatives.public.lu/initiatives/artificial-intelligence-strategic-vision-luxembourg Accessed September 23, 2024.
Government of the Republic of Estonia. (2019). Kratt – Estonian Artificial Intelligence Deployment. Retrieved from https://f98cc689-5814-47ec-86b3-db505a7c3978.filesusr.com/ugd/7df26f_27a618cb80a648c38be427194affa2f3.pdf Accessed September 23, 2024.
Government of United Kingdom. (2017). Public sector use of the cloud. Retrieved from https://www.gov.uk/guidance/public-sector-use-of-the-public-cloud Accessed September 23, 2024.
Government of United Kingdom. (2019). Understanding artificial intelligence ethics and safety. Retrieved from https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety Accessed November 4, 2022.
Government of United Kingdom. (2020a). Data Ethics framework. Government Digital Service. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/923108/Data_Ethics_Framework_2020.pdf Accessed September 23, 2024.
Government of United Kingdom. (2020b). Guidelines for AI procurement. Retrieved from https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement Accessed September 23, 2024.
Government of United Kingdom. (2020c). *Guidelines for AI procurement. A summary of best practices addressing specific challenges of acquiring Artificial Intelligence in the public sector*. UK Office for Artificial Intelligence. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/990469/Guidelines_for_AI_procurement.pdf Accessed September 23, 2024.
Government of United Kingdom. (2020d). The Government Data Quality Framework. Retrieved from https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework-guidance Accessed February 26, 2023.
Government of United Kingdom. (2020e). A guide to using artificial intelligence in the public sector. https://www.gov.uk/government/publications/a-guide-to-using-artificial-intelligence-in-the-public-sector Accessed September 23, 2024.
Government of United Kingdom. (2021a). Ethics, Transparency and Accountability Framework for Automated Decision-Making. https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making/ethics-transparency-and-accountability-framework-for-automated-decision-making Accessed September 23, 2024.
Government of United Kingdom. (2021b). Algorithmic transparency template. Retrieved from https://www.gov.uk/government/publications/algorithmic-transparency-template/algorithmic-transparency-template Accessed September 23, 2024.
Government of United Kingdom. (2021c). Algorithmic Transparency Standard. https://www.gov.uk/government/collections/algorithmic-transparency-standard Accessed December 4, 2022.
Government of United Kingdom. (2021d). Using personal data in your business or other organization. Retrieved from https://www.gov.uk/guidance/using-personal-data-in-your-business-or-other-organisation#data-protection-and-gdpr Accessed September 23, 2024.
Government of United Kingdom. (2021e). Data Protection Impact Assessments. Retrieved from https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments Accessed September 23, 2024.
Government of United Kingdom. (2022a). *Standard. Food Standards Agency*. Food Hygiene Rating Scheme – AI. Retrieved from https://www.gov.uk/government/publications/food-standards-agency-food-hygiene-rating-scheme-ai/food-standards-agency-food-hygiene-rating-scheme-ai Accessed September 23, 2024.
Government of United Kingdom. (2022b). Ethics self-assessment tool. UK statistics authority. Retrieved from https://uksa.statisticsauthority.gov.uk/the-authority-board/committees/national-statisticians-advisory-committees-and-panels/national-statisticians-data-ethics-advisory-committee/ethics-self-assessment-tool/.
Government of United Kingdom. (2022c). Equality Impact Assessment. Retrieved from https://www.gov.uk/government/consultations/emergency-evacuation-information-sharing/equality-impact-assessment Accessed September 23, 2024.
Government of United Kingdom. (2022d). Service Standard. Retrieved from https://www.gov.uk/service-manual/service-standard Accessed September 23, 2024.
Government of United Kingdom. (2022e). Data Sharing Governance Framework. Retrieved from https://www.gov.uk/government/publications/data-sharing-governance-framework/data-sharing-governance-framework Accessed February 26, 2023.
Gräf, M., Mehler, M., & Ellenrieder, S. (2024). AI strategy in action: A case study on make-or-buy for AI-based services. In *Publications of Darmstadt Technical University, Institute for Business Studies (BWL) 146709, Darmstadt Technical University, Department of Business Administration, Economics and Law*. Institute for Business Studies (BWL).
Gu, T., Dolan-Gavitt, B., & Garg, S. (2019). BadNets: Identifying vulnerabilities in machine learning model supply. *arXiv*. https://doi.org/10.48550/arXiv.1708.06733, 1708.06733v2 [cs.CR].
Gutierrez, C. I., & Marchant, G. (2021). *A global perspective of soft law programs for the governance of artificial intelligence*. Sandra Day O'Connor College of Law. Arizona State University.
Haneem, F., Kama, N., Taskin, N., Pauleen, D., & Abu Bakar, N. A. (2019). Determinants of master data management adoption by local government organizations: An empirical study. *International Journal of Information Management, 45*, 25–43. https://doi.org/10.1016/j.ijinfomgt.2018.10.007
Hernández-Nieto, R. (2002). *Contributions to statistical analysis*. Mérida: Universidad de Los Andes.
Herremans, D. (2021). aiSTROM–A roadmap for developing a successful AI strategy. IEEE. *Access, 9*, 155826–155838. https://doi.org/10.1109/ACCESS.2021.3127548
Hickman, E., & Petrin, M. (2021). Trustworthy AI and corporate governance: The EU's ethics guidelines for trustworthy artificial intelligence from a company law perspective. *European Business Organization Law Review, 22*, 593–625. https://doi.org/10.1007/s40804-021-00224-0
Hickok, M. (2022). Public procurement of artificial intelligence systems: New risks and future proofing. *AI & SOCIETY, 1-15*. https://doi.org/10.1007/s00146-022-01572-2
Holmström, J. (2022). From AI to digital transformation: The AI readiness framework. *Business Horizons, 65*(3), 329–339. https://doi.org/10.1016/j.bushor.2021.03.006
Hopster, J. (2021). What are socially disruptive technologies? *Technology in Society, 67*, Article 101750. https://doi.org/10.1016/j.techsoc.2021.101750
IEEE. (2019). Ethically aligned design. In *Committees of the IEEE global initiative on ethics of autonomous and intelligent systems. 2nd version*. Retrieved from https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf Accessed September 23, 2024.
IEEE. (2020). P7010 - Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems. Retrieved from https://standards.ieee.org/ieee/7010/7718/ Accessed September 23, 2024.
IEEE. (2021a). P7000 - Model Process for Addressing Ethical Concerns During System Design. Retrieved from https://standards.ieee.org/ieee/7000/6781/ Accessed September 23, 2024.
IEEE. (2021b). P7001 - Transparency of Autonomous Systems. Retrieved from https://standards.ieee.org/ieee/7001/6929/ Accessed September 23, 2024.
IEEE. (2021c). P7005 - Standard for Transparent Employer Data Governance Accessed September 23, 2024 https://standards.ieee.org/ieee/7005/7014/ Accessed September 23, 2024.
IEEE. (2021d). P7007 - Ontological Standard for Ethically Driven Robotics and Automation Systems. Retrieved from https://standards.ieee.org/ieee/7007/7070/ Accessed September 23, 2024.
IEEE. (2022). P7002 - Data Privacy Process. Retrieved from https://standards.ieee.org/ieee/7002/6898 Accessed September 23, 2024.
Information Commissioner''s Office. (2020). Big Data, artificial intelligence, machine learning and data protection. Version 2.2. Retrieved from https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf Accessed September 23, 2024.
Information Commissioner''s Office. (2021). International data transfers. Retrieved from https://ico.org.uk/for-organisations/dp-at-the-end-of-the-transition-period/data-protection-and-the-eu-in-detail/the-uk-gdpr/international-data-transfers/ Accessed September 23, 2024.
IPS-X. (2021). IPS-X survey. European cases. Retrieved from https://ipsoeu.github.io/ips-explorer/case/.
ISO. (2021a). ISO/IEC 24027 - Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making. Retrieved from https://www.iso.org/standard/77607.html Accessed September 23, 2024.
ISO. (2021b). ISO/IEC 24372 - Information technology — Artificial intelligence (AI) — Overview of computational approaches for AI systems. Retrieved from https://www.iso.org/standard/78508.html Accessed September 23, 2024.
ISO. (2021c). ISO/IEC 24668 - Information technology — Artificial intelligence — Process management framework for big data analytics. Retrieved from https://www.iso.org/standard/78368.html Accessed September 23, 2024.

ISO. (2022a). ISO/IEC 38507 - Information Technology – Governance implications of the use of artificial intelligence by organizations. Retrieved from https://www.iso.org/standard/56641.html Accessed September 23, 2024.

ISO. (2022b). ISO/IEC 23894 – Information Technology – Risk management. Retrieved from https://www.iso.org/standard/77304.html Accessed September 23, 2024.

IT Governance Institute. (2003). Board Briefing on IT Governance. 2nd ed. Retrieved from http://www.gti4u.es/curso/material/complementario/itgi_2003.pdf Accessed September 23, 2024.

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy artificial intelligence. *Government Information Quarterly, 37*(3), Article 101493. https://doi.org/10.1016/j.giq.2020.101493

Japanese Strategic Council for AI Technology. (2017). Artificial Intelligence Technology Strategy. Retrieved from https://ai-japan.s3-ap-northeast-1.amazonaws.com/7116/0377/5269/Artificial_Intelligence_Technology_StrategyMarch2017.pdf Accessed September 23, 2024.

Jing, H., Wei, W., Zhou, C., & He, X. (2021). An Artificial Intelligence Security Framework. In *Journal of Physics: Conference Series, Volume 1948, The 2021 2nd International Conference on Internet of Things. Artificial Intelligence and Mechanical Automation (IoTAIMA 2021), 14–16, Hangzhou, China*.

Kale, A., Nguyen, T., Harris, F. C., Li, C., Zhang, J., & Ma, X. (2022). Provenance documentation to enable explainable and trustworthy AI: A literature review. *Data Intelligence, 1-41*. https://doi.org/10.1162/dint_a_00119

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics, patterns, 2(9). *ISSN, 100314*, 2666–3899. https://doi.org/10.1016/j.patter.2021.100314

Khatri, V. (2016). Managerial work in the realm of the digital universe: The role of the data triad. *Business Horizons, 59*(6), 673–688. https://doi.org/10.1016/j.bushor.2016.06.001

Kitsios, F., & Kamariotou, M. (2021). Artificial intelligence and business strategy towards digital transformation: A research agenda. *Sustainability, 13*. https://doi.org/10.3390/su13042025

Korjani, M. M., & Mendel, J. M. (2012). Fuzzy set qualitative comparative analysis (fsQCA): Challenges and applications. In *Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS), 1-6*. https://doi.org/10.1109/NAFIPS.2012.6291026

Krippendorff, K. (2013). *Content analysis – An introduction to its methodology* (3td ed.). Sage.

Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy, 44*(6), Article 101976. https://doi.org/10.1016/j.telpol.2020.101976

Laato, S., Birkstedt, T., Mäantymäki, M., Minkkinen, M., & Mikkonen, T. (2022). AI governance in the system development life cycle: Insights on responsible machine learning engineering. In *In proceedings of the 1st international conference on AI engineering: Software engineering for AI* (pp. 113–123). https://doi.org/10.1145/3522664.3528598

Labadie, C., Legner, C., Eurich, M., & Fadler, M. (2020). FAIR enough? Enhancing the usage of Enterprise data with data catalogs. IEEE. In *22nd conference on business informatics (CBI)* (pp. 201–210). https://doi.org/10.1109/CBI49978.2020.00029

Leavy, S., O'Sullivan, B., & Siapera, E. (2020). Data, power and Bias in artificial intelligence. *ArXiv*. https://doi.org/10.48550/arXiv.2008.07341. abs/2008.07341.

Leijnen, S., Aldewereld, H., van Belkom, R., Bijvank, R., & Ossewaarde, R. (2020). *An agile framework for trustworthy AI. In NeHuAI@ ECAI* (pp. 75–78).

Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. *The Alan Turing Institute.*. https://doi.org/10.5281/zenodo.3240529

LIAA-3R. (2022). *Diretrizes de auditabilidade e conformidade no desenvolvimento e testes de soluções de IA no âmbito do LIAA-3R / Grupo de Validação Ético-Jurídica (GVEJ) do LIAA-3R, iLabTRF3, iJuspLab* (2. ed., rev. e atual). São Paulo.

Ligot, D. V.. AI governance: A framework for responsible AI development. Available at SSRN: https://ssrn.com/abstract=4817726.

Lin, Y. T., Hung, T. W., & Huang, L. T. L. (2021). Engineering equity: How AI can help reduce the harm of implicit bias. *Philosophy and Technology, 34*(Suppl. 1), 65–90. https://doi.org/10.1007/s13347-020-00406-7

Locher, M. A., & Bolander, B. (2019). Ethics in pragmatics. *Journal of Pragmatics, 145*, 83–90. ISSN 0378–2166 https://doi.org/10.1016/j.pragma.2019.01.011.

Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., & Jacquet, A. (2024). *Responsible AI pattern catalogue: A collection of best practices for AI governance and engineering. Association for Computing Machinery. New York, NY, USA* (Vol. 56, p. 7). ISSN 0360-0300. https://doi.org/10.1145/3626234

Ma, L., Zhang, Z., & Zhang, N. (2018). Ethical dilemma of artificial intelligence and its research progress. *IOP Conference Series: Materials Science and Engineering, 392*, Article 062188. https://doi.org/10.1088/1757-899X/392/6/062188

Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research, 120*, 262–273. https://doi.org/10.1016/j.jbusres.2020.07.045

Makhlouf, K., Zhioua, S., & Palamidessi, C. (2021). Machine learning fairness notions: Bridging the gap with real-world applications. *Information Processing & Management, 58*(5), Article 102642. https://doi.org/10.1016/j.ipm.2021.102642

Maluf, S. (1995). *Teoria Geral do Estado* (23ª ed., pp. 205–208). Editora Saraiva. São Paulo.

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics, 2*(4), 603–609. https://doi.org/10.1007/s43681-022-00143-x

Manzini, E. J. (2004). *Entrevista semi-estruturada: análise de objetivos e de roteiros. Seminário Internacional sobre Pesquisa e Estudos Qualitativos, 2, 2004, Bauru. A pesquisa qualitativa em debate. Anais*. Bauru: USC. isbn:85-98623-01-6.

Martin, K., & Parmar, B. (). *AI and the Creation of Knowledge Gaps: The ethics of AI transparency*. Available at SSRN: https://ssrn.com/abstract=4207128 https://doi.org/10.2139/ssrn.4207128.

McGraw, G., Bonett, R., Shepardson, V., & Figueroa, H. (2020). The top 10 risks of machine learning security. *Computer, 53*(6), 57–61. https://doi.org/10.1109/MC.2020.2984868

Medaglia, R., Gil-Garcia, J. R., & Pardo, T. A. (2021). Artificial intelligence in government: Taking stock and moving forward. *Social Science Computer Review, 1–18*. https://doi.org/10.1177/08944393211034087

Meijerink, J., & Bondarouk, T. (2018). Uncovering configurations of HRM service provider intellectual capital and worker human capital for creating high HRM service value using fsQCA. *Journal of Business Research, 82*, 31–45. https://doi.org/10.1016/j.jbusres.2017.08.028

Mezgár, I., & Váncza, J. (2022). From ethics to standards – A path via responsible AI to cyber-physical production systems. *Annual Reviews in Control, 53*, 391–404. ISSN 1367-5788. https://doi.org/10.1016/j.arcontrol.2022.04.002

Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society, 7*(2). https://doi.org/10.1177/2053951720948087

Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side' of AI. *European Journal of Information Systems, 31*(3), 257–268. https://doi.org/10.1080/0960085X.2022.2026621

Mikalef, P., Lemmer, K., Schaefer, C., Ylinen, M., Fjørtoft, S. O., Torvatn, H. Y., … Niehaves, B. (2022). Enabling AI capabilities in government agencies: A study of determinants for European municipalities. *Government Information Quarterly, 39*(4), Article 101596.

Ministério da Ciência Tecnologia e Inovação do Brasil. (2021). Estratégia Brasileira de Inteligência Artificial. Governo do Brasil. Retrieved from https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ia_estrategia_documento_referencia_4-979_2021.pdf Accessed September 23, 2024.

Ministero dello sviluppo economico. (2019). Proposte per uma Strategia italiana per l'intelligenza artificiale. Retrieved from https://www.mise.gov.it/images/stories/documenti/Proposte_per_una_Strategia_italiana_AI.pdf.

Ministry of Economic Affairs and Employment of Finland. (2017). Suomen tekoälyaika. Retrieved from https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y Accessed September 23, 2024.

Ministry of Economic Affairs and Employment of Finland. (2019). *Leading the Way into the Era of Artificial Intelligence: Final Report of Finland's Artificial Intelligence Program 2019. Ministry of Economic Affairs and Employment of Finland* (p. 133). Retrieved from http://urn.fi/URN:ISBN:978-952-327-437-2 Accessed September 23, 2024.

Misuraca, G., & van Noordt, C. (2020). *Overview of the use and impact of AI in public services in the EU, EUR 30255 EN. 2020*. Luxembourg: Publications Office of the European Union. ISBN 978–92–76-19540-5, doi:10.2760/039619, JRC120399.

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: an initial review of publicly available ai ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics, 26*, 2141–2168. https://doi.org/10.1007/s11948–019-00165-5

MRE. (2022). De outros países no Brasil. Retrieved from https://www.gov.br/mre/pt-br/assuntos/Embaixadas-Consulados-Missoes/de-outros-paises-no-brasil Accessed September 23, 2024.

Nagbøl, P. R., & Müller, O. (2020). X-RAI: A framework for the transparent, responsible, and accurate use of machine learning in the public sector. In *IFIP EGOV-ePart-CeDEM conference. p. 259. CEUR workshop proceedings*.

Nagbøl, P. R., Müller, O., & Krancher, O. (2021). Designing a risk assessment tool for artificial intelligence systems. In the next wave of sociotechnical design. In *, 16. 16th international conference on design science research in information systems and technology, DESRIST 2021, Kristiansand, Norway, august 4–6, 2021, proceedings* (pp. 328–339). Springer International Publishing. https://doi.org/10.1007/978-3-030-82405-1_32.

National Institute of Advanced Industrial Science and Technology. (2022). Machine Learning Quality Management Guideline – 2nd English edition. Government of Japan – Digital Architecture Research Center. Retrieved from https://www.digiarc.aist.go.jp/en/publication/aiqm/aiqm-guideline-en-2.1.1.0057-e26-signed.pdf Accessed September 23, 2024.

Navarro, S., Llinares, C., & Garzon, D. (2016). Exploring the relationship between cocreation and satisfaction using QCA. *Journal of Business Research, 69*(4), 1336–1339.

NIST. (2022). AI Risk Management Framework: first draft. Retrieved from https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf Accessed September 23, 2024.

Nordic Council of Ministers. (2018). AI in the Nordic-Baltic region. Retrieved from https://www.norden.org/en/declaration/ai-nordic-baltic-region Accessed September 23, 2024.

Norwegian Data Protection Authority. (2018). Datatilsynet. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf Accessed September 23, 2024.

Norwegian Ministry of Local Government and Modernisation. (2020). National Strategy for Artificial Intelligence. Retrieved from https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/en-gb/pdfs/ki-strategi_en.pdf Accessed September 23, 2024.

Ntoutsi, E., Fafaliosb, P., Gadirajua, U., Iosidisa, V., Nejdla, W., Vidalc, M., … Staab, S. (2020). Bias in Dat-driven AI systems – An introductory survey. *arXiv*. https://doi.org/10.48550/arXiv.2001.09762, 2001.09762v1 [cs.CY].

OECD. (2022a). Artificial Intelligence Observatory. Retrieved from https://oecd.ai/en/ Accessed September 23, 2024.
OECD. (2022b). United States – Local state and Federal Regulations on facial recognition technologies. Organization for Economic and co-operation Development Artificial Intelligence Observatory. Retrieved from retrieved from https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-26890.
OECD. (2022c). Policy Initiatives for Emerging AI-related regulation, Civil society. https://oecd.ai/en/dashboards/policy-initiatives?conceptUris=http:%2F%2Fai.oecd.org%2Fmodel%23Emerging_technology_regulation%7C%7Chttp:%2F%2Fai.oecd.org%2Ftaxonomy%2FtargetGroups%23TG16 Accessed September 23, 2024.
OECD/CAF. (2022). *The strategic and responsible use of artificial intelligence in the public sector of Latin America and the Caribbean*. Paris: OECD Public Governance Reviews, OECD Publishing. https://doi.org/10.1787/1f334543-en
Ojo, A., Mellouli, S., & Ahmadi Zeleti, F. (2019). A realist perspective on AI-era public management. In *In 20th annual international conference on digital government research* (pp. 159–170). ACM.
Oneto, L., & Chiappa, S. (2020). *Fairness in machine learning. arXiv:2012.15816 [cs.LG]*.
Özdemir, V., & Hekim, N. (2018). Birth of industry 5.0: Making sense of big data with artificial intelligence,"the internet of things" and next-generation technology policy. *OMICS: A Journal of Integrative Biology, 22*(1), 65–76. https://doi.org/10.1089/omi.2017.0194
Papagiannidis, E., Enholm, I. M., Dremel, C., Mikalef, P., & Krogstie, J. (2023). Toward AI governance: Identifying best practices and potential barriers and outcomes. *Information Systems Frontiers, 25*, 123–141. https://doi.org/10.1007/s10796-022-10251-y
Phillips, P. J., Hanan, C. A., Fontana, P. C., Broniatowski, D. A., & Przybocki, M. A. (2021). *Four principles of explainable artificial intelligence*. National Institute of Standards and Technology. U.S. Department of Commerce.
Pinski, M., Hofmann, T., & Benlian, A. (2024). AI literacy for the top management: An upper echelons perspective on corporate AI orientation and implementation ability. *Electronic Markets, 34*, 24. https://doi.org/10.1007/s12525-024-00707-1
Presidencia de la Nación. (2019). ARGENIA – Plan Nacional de Inteligencia Artificial. Argentina. Retrieved from https://oecd-opsi.org/wp-content/uploads/2021/02/Argentina-National-AI-Strategy.pdf Accessed September 23, 2024.
Ragin, C. C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond* (pp. 85–97). Chicago: Univ. of Chicago Press.
Rahwan, I. (2017). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology, Springer., 20*, 5–14. https://doi.org/10.1007/s10676-017-9430-8
Rahwan, I., Cebrian, M., Obradovich, N., et al. (2019). Machine behaviour. *Nature, 568*, 477–486. https://doi.org/10.1038/s41586-019-1138-y
Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine, 169*(12), 866–872. https://doi.org/10.7326/M18-1990
Rhahla, M., Allegue, S., & Abdellatif, T. (2021). Guidelines for GDPR compliance in big data systems. *Journal of Information Security and Applications, 61*, Article 102896. https://doi.org/10.1016/j.jisa.2021.102896
Rihoux, B., & Ragin, C. (2008). *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. London and Thousand Oaks, CA: Sage.
Roorda, S. A. H. (2021). *Facial recognition for public safety a supportive tool for the municipal decision-making process on using facial recognition for public safety, the FRPS risk governance method*. Master's thesis,. University of Twente.
Rose, J., Flack, L. S., & Sæbø, Ø. (2018). Stakeholder theory for the E-government context: Framing a value-oriented normative core. *Government Information Quarterly, 35*(3), 362–374. ISSN 0740-624X https://doi.org/10.1016/j.giq.2018.06.005.
Roselli, D., Matthews, J., & Talagala, N. (2019). *Managing Bias in AI. In Companion Proceedings of The 2019 World wide web conference (WWW '19)* (pp. 539–544). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3308560.3317590
Ruijer, E. (2021). Designing and implementing data collaboratives: A governance perspective. *Government Information Quarterly, 38*(4), Article 101612. https://doi.org/10.1016/j.giq.2021.101612
Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). Sage.
Schaefer, C., Lemmer, K., Samy Kret, K., Ylinen, M., Mikalef, P., & Niehaves, B. (2021). Truth or dare?–how can we influence the adoption of artificial intelligence in municipalities?. Retrieved from http://hdl.handle.net/10125/70899.
Schrader, D., & Ghosh, D. (2018). Proactively protecting against the singularity: Ethical decision making AI. *IEEE Computer and Reliability Societies Review, 16*(3), 56–63.
Schüller, K. (2022). Data and AI literacy for everyone. *Jan, 1*, 477–490. https://doi.org/10.3233/sji-220941
Shao, S., Zhao, R., Yuan, S., Dig, M., & Wang, Y. (2022). Tracing the evolution of AI in the past decade and forecasting the emerging trends. *Expert Systems with Applications., 209*, Article 118221. https://doi.org/10.1016/j.eswa.2022.118221
Sharma, G. D., Yadav, A., & Chopra, R. (2020). Artificial intelligence and effective governance: A review, critique and research agenda. *Sustainable Futures, 2*. https://doi.org/10.1016/j.sftr.2019.100004. ISSN 2666–1888.
Sigfrids, A., Leikas, J., Salo-Pöntinen, H., & Koskimies, E. (2023). Human-centricity in AI governance: A systemic approach. *Frontiers in Artificial Intelligence, 6*, Article 976887. https://doi.org/10.3389/frai.2023.976887. PMID: 36872934; PMCID: PMC9979257
Silberg, J., & Manyika, J. (2019). *Notes from the AI frontier: Tackling bias in AI (and in humans). 1(6)*. McKinsey Global Institute. Retrieved from https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/tackling%20bias%20in%20artificial%20intelligence%20and%20in%20humans/mgi-tackling-bias-in-ai-june-2019.pdf Accessed September 23, 2024.
Silveira, M. B., Saldanha, R. P., Leite, J. C. C., Silva, T. O. F. D., Silva, T., & Filippin, L. I. (2018). Construction and validation of content of one instrument to assess falls in the elderly. *einstein (Sao Paulo), 11;16*(2), Article eAO4154. Retrieved from https://doi.org/10.1590/S1679-45082018AO4154 accessed September 23, 2024.
Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research, 70*, 263–286.
Smuha, N. A. (2021). Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea. *Philos. Technol., 34*(Suppl 1), 91–104. https://doi.org/10.1007/s13347-020-00403-w
Stahl, B. C., Antoniou, J., Ryan, M., et al. (2022). Organisational responses to the ethical issues of artificial intelligence. *AI & SOCIETY, 37*, 23–37. https://doi.org/10.1007/s00146-021-01148-6
Stix, C. (2021). Foundations for the future: Institution building for the purpose of artificial intelligence governance. *AI and Ethics*. https://doi.org/10.1007/s43681-021-00093-w
Strauß, S. (2021). Deep automation Bias: How to tackle a wicked problem of AI? *Big Data Cogn. Comput., 5*, 18. https://doi.org/10.3390/bdcc5020018
Switzerland Federal Council. (2021). Guidelines on Artificial Intelligence for the Confederation. Retrieved from https://www.sbfi.admin.ch/dam/sbfi/en/dokumente/2021/05/leitlinien-ki.pdf.download.pdf/leitlinien-ki_e.pdf Accessed September 23, 2024.
Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society, 40*(2), 137–157. https://doi.org/10.1080/14494035.2021.1928377
Tangi, L., van Noordt, C., Combetto, M., Gattwinkel, D., & Pignatelli, F. (2022). *AI watch. European landscape on the use of artificial intelligence by the public sector, EUR 31088 EN*. Luxembourg: Publications Office of the European Union. ISBN 978-92-76-53058-9, doi:10.2760/39336, JRC129301.
Torlig, E. G. S., Resende-Júnior, P. C., & Fujihara, R. K. (2019). Proposição de uma Nova Orientação para Validação de Roteiros em Pesquisas Qualitativas. In *XLIII Encontro da ANPAD – EnANPAD 2019, São Paulo*.
Torlig, E. G. S., Resende-Júnior, P. C., Fujihara, R. K., Demo, G., & Montezano, L. (2022). Validation Proposal for Qualitative Research Scripts (Vali-Quali). *Administração: Ensino e Pesquisa, 23*(1). https://doi.org/10.13058/raep.2022.v23n1.2022, 4-29; Jan-Abr.
Vero. (2019). Finnish Tax Administration's ethical principles for AI. Retrieved from https://www.vero.fi/en/About-us/finnish-tax-administration/operations/responsibility/finnish-tax-administrations-ethical-principles-for-ai/#:~:text=Our%20AI%20follows%20laws%20and%20regulations&text=The%20use%20of%20AI%20does,our%20partners%20carefully%20and%20responsibly.
Vetrò, A., Torchiano, M., & Mecati, M. (2021). A data quality approach to the identification of discrimination risk in automated decision-making systems. *Government Information Quarterly, 38*(4). https://doi.org/10.1016/j.giq.2021.101619. ISSN 0740-624X.
Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems, 28*(2), 118–144. ISSN 0963-8687. https://doi.org/10.1016/j.jsis.2019.01.003
Vilminko-Heikkinen, R., & Pekkola, S. (2019). Changes in roles, responsibilities and ownership in organizing master data management. *International Journal of Information Management, 47*, 76–87. ISSN 0268-4012, https://doi.org/10.1016/j.ijinfomgt.2018.12.017.
Vining, R., McDonald, N., McKenna, L., Ward, M. E., Doyle, B., Liang, J., … Fogarty & Brennan, R. (2022). Developing a Framework for Trustworthy AI-Supported Knowledge Management in the Governance of Risk and Change. In *HCI International 2022-Late Breaking Papers. Design, User Experience and Interaction: 24th International Conference on Human-Computer Interaction, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings* (pp. 318–333). Cham: Springer International Publishing.
WEF. (2020). Unblocking Public Sector AI – AI procurement in a Box: Project overview. Retrieved from https://mkto.deloitte.com/rs/712-CNF-326/images/Retningslinjer-for-offentlige-AI-anskaffelser.pdf Accessed September 23, 2024.
WEF. (2020b). Unblocking Public Sector AI. AI Procurement in a Box: Pilot case studies from the United Kingdom. Retrieved from https://www.weforum.org/reports/ai-procurement-in-a-box/#case-study-uk Accessed September 23, 2024.
Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2018). Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration, 42*(7), 596–615.
Wirtz, B. W., Weyerer, J. C., & Kehl, I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. *Government Information Quarterly*. https://doi.org/10.1016/j.giq.2022.101685. ISSN 0740-624X.
Wright, S. A., & Schultz, A. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. *Business Horizons, 61*(6), 823–832. https://doi.org/10.1016/j.bushor.2018.07.001
Xue, M., Yuan, C., Wu, H., Zhang, Y., & Liu, W. (2020). Machine learning security: Threats, countermeasures, and evaluations. *IEEE Access, 8*(74720–74742), 2020. https://doi.org/10.1109/ACCESS.2020.2987435
Zhou, J., & Chen, F. (2023). AI ethics: From principles to practice. *AI & SOCIETY, 38*, 2693–2703. https://doi.org/10.1007/s00146-022-01602-z
Zicari, R. V., Brodersen, J., Brusseau, J., Düdder, B., Eichhorn, T., Ivanov, T., … Westerlund, M. (2021). Z-inspection®: A process to assess trustworthy AI. *IEEE Transactions on Technology and Society, 2*(2), 83–97. https://doi.org/10.1109/TTS.2021.3066209
Zick, T., Kortz, M., Eaves, D., & Doschi-Velez, F. (2024). *AI Procurement Checklists: Revisiting Implementation in the Age of AI Governance*. arXiv:2404.14660v1 [cs.CY]. https://doi.org/10.48550/arXiv.2404.14660.
Zuiderwijk, A., Chen, Y., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research

agenda. *Government Information Quarterly, 38*(3). https://doi.org/10.1016/j.giq.2021.101577. ISSN 0740-624X.

Patricia Gomes Rêgo de Almeida has Master degree at University of Rio Grande do Norte, Department of Electrical Engineering, and PhD degree at the University of Brasilia (UnB), Department of Business Administration, where she has developed research regarding artificial intelligence regulation (hard and soft law) and artificial intelligence governance. She is member of Linselab at University of Brasilia (UnB). She is the coordinator of Digital Innovation, Governance and Strategy at the Brazilian chamber of Deputies, where she is responsible for its Digital Transformation Strategy, Artificial Intelligence Governance and Data Fluency Program. She is coordinator of the Parliamentary Data Science Hub at the Inter-parliamentary Union, where she coordinated the writing of Guidelines for AI in Parliaments.

Carlos Denner dos Santos Júnior Junior is professor at University of Brasilia (UnB), Department of Business Administration, focusing on the intersection between administration and computing, developing cutting-edge scientific knowledge about the creation and management of new businesses and information technologies with applications in public administration and their impacts on the market in general. He has post-doctoral degrees at: Université du Quebec à Montréal (UQUAM), Universidade Federal de Pernambuco (UFPE), University of Nottingham, and Universidade de São Paulo (USP). He is the coordinator of the Research Group at CNPq Sociedados - on the Strategic and Competitive Use of Data (Open) and Software (Free), and the DINTER coordinator between UnB-UNIMONTES.

Publication 2 : Artificial Intelligence Regulation

Artificial Intelligence Regulation: a framework for governance

Article *in* Ethics and Information Technology · September 2021

DOI: 10.1007/s10676-021-09593-z

3 authors:

Patricia Gomes Rêgo de Almeida
Chamber of Deputies (Brazil)
6 PUBLICATIONS **468** CITATIONS
SEE PROFILE

Carlos Denner dos Santos Jr.
University of Brasília
64 PUBLICATIONS **880** CITATIONS
SEE PROFILE

Josivania Silva Farias
University of Brasília
61 PUBLICATIONS **1,067** CITATIONS
SEE PROFILE

Ethics and Information Technology
https://doi.org/10.1007/s10676-021-09593-z

ORIGINAL PAPER

Artificial Intelligence Regulation: a framework for governance

Patricia Gomes Rêgo de Almeida[1,2] · **Carlos Denner dos Santos**[1,3] · **Josivania Silva Farias**[1]

Accepted: 7 April 2021

Abstract
This article develops a conceptual framework for regulating Artificial Intelligence (AI) that encompasses all stages of modern public policy-making, from the basics to a sustainable governance. Based on a vast systematic review of the literature on Artificial Intelligence Regulation (AIR) published between 2010 and 2020, a dispersed body of knowledge loosely centred around the "framework" concept was organised, described, and pictured for better understanding. The resulting integrative framework encapsulates 21 prior depictions of the policy-making process, aiming to achieve gold-standard societal values, such as fairness, freedom and long-term sustainability. This challenge of integrating the AIR literature was matched by the identification of a structural common ground among different approaches. The AIR framework results from an effort to identify and later analytically deduce synthetic, and generic tool for a country-specific, stakeholder-aware analysis of AIR matters. Theories and principles as diverse as Agile and Ethics were combined in the "AIR framework", which provides a conceptual lens for societies to think collectively and make informed policy decisions related to what, when, and how the uses and applications of AI should be regulated. Moreover, the AIR framework serves as a theoretically sound starting point for endeavours related to AI regulation, from legislation to research and development. As we know, the (potential) impacts of AI on society are immense, and therefore the discourses, social negotiations, and applications of this technology should be guided by common grounds based on contemporary governance techniques, and social values legitimated via dialogue and scientific research.

Keywords Ethics · Artificial Intelligence · Regulation · Governance · Framework

Introduction

The widespread use of AI in our daily actions and in an unnoticeable fashion (Cerka et al., 2015) has introduced unprecedented ethical issues to a broad and complex social system (Cave et al., 2019).

From the same perspective, the complexity of data treatment in the design and development process of a machine learning solution increases the likelihood of ethical surprises, which demands a wider evaluation of the ethical and social impacts (Butterworth, 2018).

Based on this reflection, this work has sought to conduct a vast search for literature that is relevant in terms of Artificial Intelligence Regulation, processing and grouping it into a set of purposes presented as frameworks or guidelines for a framework based on ethical principles. Their main contributions have been customised as a framework based on the Design and Action Theory (Gregor, 2006) that allows for reflections and actions aimed at regulating and governing operations and relationships between natural and legal persons on one side, and AI-embedded systems on the other.

✉ Patricia Gomes Rêgo de Almeida
patricia.almeida@camara.leg.br

Carlos Denner dos Santos
carlosdenner@unb.br

Josivania Silva Farias
josivania@unb.br

[1] University of Brasilia (UnB) – Department of Administration, Brasília, Brazil

[2] Chamber of Deputies of Brazil – Directorate of Innovation and Information Technology, Brasília, Brazil

[3] LATECE, University of Quebec at Montreal (UQAM), Montreal, Canada

Published online: 21 April 2021

Reasons to regulate AI

Since the term was coined in 1956, Artificial Intelligence has been associated with a wide range of concepts (Cerka et al., 2017; Jackson, 2019) based on a thinking human being and on rational behaviour, which could be synthetised as: systems that think and act like humans and systems that think and act rationally (Cerka et al., 2015; Russell & Norvig, 1995). Equally wide is the variety of different names associated with whatever utilises AI technology: robots, smart systems, intelligent systems, intelligent agents, AI agents, AI algorithms, intelligent algorithms, and autonomous systems, to mention a few.

For the purpose of avoiding misunderstandings regarding AI, the High-Level Expert Group established by the European Commission has defined AI systems as "software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions." (AI HLEG, 2019a). Considering the difficulty of defining AI in a way that could fit all approaches needed for regulation and governance actions with clear communication among all stakeholders, this article adopts the definition of AI established by the High-Level Expert Group on AI systems.

The responsibilities, security, intellectual property, and privacy associated with different systems for medical robots, drones, autonomous cars, among several "intelligent solutions" offered every day have been questioned.

Illustrating the level of risk-related indeterminacy, machine learning has been combined with game theory (Conitzer et al., 2017) in cases where developers were using game theory to help teach strategic defence to algorithms. A game between two algorithms predicted that one would kill the other only in case of an absolute scarcity of resources. However, when a more intelligent algorithm was introduced, it immediately killed the weaker ones (Firth-Butterfield, 2017). This case reinforces the idea that an autonomous system will inevitably find itself in a situation in which it needs not only to obey a certain rule or not, but also to make a complex ethical decision (Dennis et al., 2016).

Facing the risks compels us to explore their causes and effects. Although the effects of AI are not yet known, a large amount of them can currently be classified. Firstly, those coming from the undesired effects, such as biases, discrimination, loss of privacy, false positives and false negatives, loss of autonomy, (psychological, financial, or physical) damage, loss of control, difficulty identifying liabilities, losses or decreases in human rights, unemployment, misjudgements, and concentration of power and wealth in a few companies. Secondly, some risks are the result of intentional misuses, such as fake news, deep fakes, cyberattacks, terrorism, warfare, weapons, people manipulation, espionage, low level of democracy (Beltran, 2020; Benjamins & Garcia 2020; Borgesius, 2018; Jackson, 2020; Jobin et al., 2019; Mika et al., 2019).

Considering all those risks, establishing best practices for delegating and defining new moral responsibility attribution models is crucial to leverage the opportunities created by AI (Taddeo & Floridi, 2018). Risk assessment models can provide support and flexibility for Big Data and AI applications (Mantelero, 2018), and stakeholders who develop and deploy AI-based systems must enhance their knowledge of the values protected by human rights and how those rights apply to their own actions (Smuha, 2020).

Despite being a huge challenge, finding a way to deal with ethical issues must be a constant target of research, for what we need to join all our forces (Bostrom, 2014), and AI regulation is on the right path to get there (Carter, 2020).

The reasons to regulate include: manufacturers' need to comprehend a legal framework within which they can operate reliably; consumers' and society's need to be protected from devices that may harm or adversely affect them; and the need for business opportunities (Holder et al., 2016a).

In industries still lacking regulation, the general approach observed is that innovation is freely allowed, but those in charge should bear the consequences in case certain types of damage are caused (Reed, 2018).

Faced with the challenge of minimising those risks, a combination of strategy and actions must be put to practice during the entire lifecycle of AI systems, in order not only to identify damages and responsibilities, but also, and especially, to avoid them.

Seeking the best way to regulate

Sometimes, when used to denote an attempt to standardise behavioural patterns, the term "regulation" assumes the meaning of a law (Hildebrandt, 2018).

However, on a broader approach, regulation is a sustained attempt to modify behaviours of others according to defined standards or purposes in order to produce the desired outcomes. This can involve standard-setting, information-gathering, and behaviour modification mechanisms (Black, 2002), especially in cases evolving ethical issues, whose understanding is complex when applied to a real world. Therefore, law is just one way of regulating society, while

other alternatives to regulate human behaviour may also be widely used (Hildebrandt, 2018).

Disruptive innovation always challenges regulatory strategies due to the reactive nature of traditional regulation (Kaal & Vermeulen, 2017). In the case of innovation by AI, the challenge is amplified, since it is strongly related to ethical issues and its results could be unpredictable in some situations, bringing about unforeseen social impacts. In addition, if AI adoption and implementation are conducted in a reckless manner, social and political instability could ensue, thus threatening freedom, self-determination, human rights, and fundamental values (Caron & Gupta, 2020). As human behaviour encompasses decisions from an ethical perspective, the regulation should also consider it. While norms as instruments of regulation relate to what is good or bad from society's point of view, ethics concerns itself with the nature of the principles upon which those norms are founded (Pedro, 2014).

A few laws have been resorted to in an attempt to settle damages caused by AI-supported products and services judicially. If, on the one hand, the number of cases is multiplying, on the other, the legislative branch seems to be moving at a negligible speed compared to the technological advancements enforcing the perception that traditional regulation does not fit in this challenge (Cerka et al., 2015; Larsson, 2020; Villaronga & Heldeweg, 2018). Part of this increasing gap between laws and technology is caused by the lack of a thorough and accurate definition of AI (Firth-Butterfield, 2017; Larsson, 2020), which is aggravated by the fact that the definition changes as the technology evolves (Fjeld et al., 2020). Considering this issue, the concept of dynamic regulation could fit in the field of AI, as it is based on learning by doing and continuity of regulatory relationships (Kaal & Vermeulen, 2017; Lewis & Yildirim, 2002).

A yet-to-be-solved equation is the breadth of laws dealing with globally produced and commercialised technologies (Holder et al., 2016a) and robot-generated inventions (Holder et al., 2016b). The problem reaches even broader dimensions when one considers the complex networks established in the technology industry, making it possible for products to be subjected to learning from data scattered across the world (Lenardon, 2017).

Large-scale data analyses have revealed that the key challenge related to the AI regulation dilemma is demonstrating it is produced and deployed appropriately (Butterworth, 2018). One of the most advocated strategies is transparency, an opening of the entire production process, especially the decision-making rules, the method, and the data utilised when training the intelligent system (Buiten, 2019; Butterworth, 2018; Tutt, 2017). However, on certain occasions, even in case the AI algorithm is open, full transparency cannot be ensured, as there is a difference between seeing the whole code and understanding all of its potential effects (Firth-Butterfield, 2017). A similar strategy to open data is the Explainable Artificial Intelligence (XAI) standard for the creation of coding models oriented towards a global comprehension (Adadi & Berrada, 2018; Taddeo & Floridi, 2018). In addition to the concerns related to the development process of an AI system, data governance has been recognised as being key to AI governance (Hilb, 2020; UK Government, 2018).

Some of the AI regulation theories that have been proposed are based on contractual and extracontractual liability, or on strict liability, and adopt a liability model in which the moral responsibility is distributed among designers, regulators, and users. The attempt to hold robots accountable for their actions has led a few countries to consider the possibility of granting a legal identity to each unit. One could argue that if parties in a contractual relationship may be legally represented by another entity, then so can systems (Cerka et al., 2017). As a counterargument, the term "robot liability" should be replaced with "indirect liability over the robot", given the impossibility of claiming damages from a robot, i.e., it cannot be held criminally liable. Thus, the impact of such products on society should also be a liability (Jackson, 2019; Nevejans, 2016). Although this latter understanding tends to be more acceptable from a global perspective, a liability model is still an essential and complex variable to be defined through an AI regulation strategy.

Also among the concerns that motivate AI regulation is the approach aimed at minimising the disruption of the work model with the goal of fighting job loss (Wright & Schultz, 2018).

Drawing attention to the domain of what is to be regulated, attempts to legislate digital technologies without proper knowledge for doing so have been criticised (Reed, 2018). With the intention of minimising those risks, a gradual regulation strategy (Villaronga & Heldeweg, 2018) can be used. When mitigating risks, regulatory agencies could bar the introduction of certain algorithms into the market until their safety and efficacy have been proven by means of tests (Tutt, 2017) founded on ethics (Arkin, 2011).

In 2017, the European Parliament Committee on Legal Affairs released a report recommending the creation of a European agency for robotics and AI, suggesting a combination of both hard and soft laws, given the complexity associated with the evolution of the regulatory model. It would put regulators and external experts together to monitor AI trends and study standards for best practices (Cath et al., 2017; Nevejans, 2016). After approving the study of the High-Level Expert Group on AI, the European Commission recommended upgrading the European Framework to one especially designed for AI Governance (European Commission, 2019). In the same direction, the House of Lords (2018) has recommended the creation of an AI regulatory framework.

Another effort observed in the US has resulted in S.3891, which defines conditions for advancing Artificial Intelligence research, including the development of technical standards (US Congress, 2020), and in H.Res. 153, which aims to support the development of guidelines for the ethical dsevelopment of Artificial Intelligence (US Congress, 2019).

In a parallel effort, many self-regulatory private-sector initiatives have been created, and research has been carried out to discuss ethical issues on AI development and use, such as the Partnership on AI to Benefit People and Society (AI4People, 2018; Partnership on AI, 2016; The Future of Life Institute, 2019b), The Montreal Declaration for a Responsible Development of Artificial Intelligence (University of Montreal, 2018), and The Toronto Declaration (Toronto, 2020).

At the government level, ethical principles were considered when the national AI-oriented strategies of a few countries were drawn up, as happened in Japan (Japanese Cabinet Office, 2019), France (French PM, 2018), Germany (German Federal Government, 2018), United Arab Emirates (Dubai, 2019), India (Aayog, 2018), and Singapore (Monetary Authority of Singapore, 2019). Additionally, several countries have shown their intention to create policies and laws to regulate the development and use of AI (Future of Life Institute, 2019a). Similar concerns have served as the basis for recommendations regarding ethical principles by a few transnational organisations, such as the Council of Europe (2018) and the Organisation for Economic Cooperation and Development (2019).

As the major concern regarding both self-regulation and government initiatives kickstarted the debate on AI governance through ethical principles, a set of core topics was comprised in each one of them: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values. However, different principles can be observed under the same topic, which illustrates the lack of unanimity (Fjeld et al., 2020).

Standardisation documents are also part of the efforts associated with the AI regulation challenge. A good example is the ad hoc technical committee on Artificial Intelligence established within the International Organisation for Standardisation (ISO), whose plan includes two dozen standards on AI and Big Data (Neznamov, 2020).

The huge gap between ethical guidelines and laws, apart from the great number of potential situations in which stakeholders (developers, deployers, etc.) fail to apply such ethical principles, draws attention to the need to shift from principles to processes when it comes to AI governance (Larsson, 2020). Thus, there is a long path to be paved through a connected network of processes including several stakeholders in a way to keep on pace with the society's values.

Method

With the goal of surveying the relevant scientific literature on AI regulation, we have systematically searched for and organised papers to summarise the corpus and perform a qualitative analysis to understand the evolution and current state of the science.

We have compiled papers published between 2010 and 2020 containing the following expressions: ("ARTIFICIAL INTELLIGENCE" and "ETHICAL USE"), ("ARTIFICIAL INTELLIGENCE" and "REGULATION"), or ("ARTIFICIAL INTELLIGENCE" and "GOVERNANCE"). This search then resulted in title and subject matches on the ScienceDirect, JSTOR, SpringerLink, PROQUEST, IEEE, Scopus, DOAJ, and Google Scholar databases. Only peer-reviewed research articles in English have been compiled.

The selection of papers was later refined by reading all abstracts with the goal of removing case-specific discussions, as well as those in which regulation was not the main topic under debate. In addition, new papers on AI-related laws and government strategies that presented arguments on ethical principles have also been included in the sample.

This final corpus of literature has been classified according to specific parameters: year of publication, journal, author, author's institution, author's field of study, country, keywords. Summaries of each paper have also been developed to include: concepts, findings, contributions, agenda, approach, method, and researched subject. The following terms were considered when classifying the articles: "ethics/ethical principles", "how to regulate/existing regulation", "government strategies", and "framework or guidelines similar to a framework". After analysing the abstracts, a sample comprising 109 documents was selected for further reading and inclusion.

Results

In chronological terms, it is worth highlighting that 88.1% of the papers were published after 2015, with a growing production every year following that.

The sample reflects the evolution in the fields of research that take an interest in AI regulation, which is desired (Floridi et al., 2020). Although Artificial Intelligence as a subject of study traditionally pertains to Information Technology (Computer Science and Engineering), there has been a growing interest in its regulation by other areas, such as Law, Business Administration, and Philosophy. Out of the entire sample, researchers from the field of IT (combined or not with other fields) represent 47.7% and researchers exclusively from IT represent 27.5%, whereas researchers from other fields (excluding IT) represent 52.3%. In some

cases, the same article is co-authored by researchers from different areas (22%).

Special heed has been paid to the analysis of the main object of the sample, non-exclusively divided into: "ethics and ethical principles" (45.9%), "how to regulate and existing regulation" (47.7%), "government strategies" (9.2%) and "framework or guidelines similar to a framework" (19.3%). It is worth noting that discussions on how to regulate only became significant in 2016. Concerning the discussions on AI regulatory frameworks, AI guidelines based on ethical principles with orientations similar to a framework have also been considered when they go beyond a description of ethical principles and actually provide orientations concerning how to apply those principles. From that perspective, 21 unique models have been found, which will be presented and analysed below.

Model for ethical issues in experimental technologies (Amigoni & Schiaffonati, 2018)

Based on the premise that a robot is an experimental technology, this model intends to minimise the ethical dilemmas associated with decisions made by autonomous systems (Poel, 2016). The proposal supports decision-making processes based on 16 conditions for deploying experimental technologies built to anticipate potential ethical issues as robots interact with people and the environment. Split into three groups, the conditions are aimed at: preventing damages (non-maleficence conditions: means for gaining knowledge of risks and benefits, monitoring of data and risks, possibility and willingness to adapt or terminate the experiment, risk mitigation, consciously scaling up, flexible setup, and avoidance of locking-in and undermining resilience); good-doing (beneficence conditions: expectation of social benefits, clear distribution of responsibilities); and respect for autonomy and justice (experimental subjects informed, approval by democratically legitimised bodies, possibility of experimental subjects influencing the project, possibility of withdrawing subjects from the experiment, special treatment given to vulnerable experimental subjects, fair distribution of potential hazards and benefits, reversibility of compensation of harm).

The model is an approach to regulation through a development process that would be part of a gradual interactive strategy set forth during the design stage. One can find among the outputs the epistemological role of exploratory experiments, while acquiring the knowledge of how robots behave in a real-world scenario. The authors highlight the prediction of "red button" conditions for situations in which the risk of harming people cannot be securely avoided during the experiment.

The 16 conditions proposed by Amigoni and Schiaffonati's "Ethical Framework for Robot Systems" seem to fit perfectly in standardised processes built by regulatory agencies as they test all the technologies submitted by the industry and service providers. The proposal can also be incorporated through risk analyses conducted by scholars for society as a whole.

Interactive regulatory governance model (Villaronga & Heldeweg, 2018)

Considering that regulatory actions cannot keep up with the speed of technology, and that top-down regulation approaches require mature laws, the authors have identified the need for a hybrid approach to start regulating AI technologies. They argue that bottom-up mechanisms can help develop the legislation and produce knowledge of AI development processes.

Focusing on a balance between regulation/legislation-in-progress and technology-in-progress, the proposal is based on an interactive governance model for technological development and law formulation processes in which the attributions of stakeholders are highlighted through process descriptions. The need for continuous learning and a gradual evolution of the legal framework is noteworthy, using such expressions as "Regulatory Innovation" and "Temporary Experimental Legislation", and considering the proper sequence of actions among agents at the maturity stage of an innovation's lifecycle.

The proposed model includes components such as:

- A Regulatory-to-Technology (R2T) macro-process to guide the creation of a new conceptual model for robots in accordance with the existing legislation, considering how it affects the way intelligent systems are built and used. It enables the creation of an AI technology impact assessment encompassing ethical, legal, and societal consequences. It focuses on legal opportunities or constraints that could have an impact on a new or existing robot. The result of the analysis considers a range of alternatives, from "abort development", "adjust plans", "go-ahead and lobby for legal change", or "take risks".
- A Technology-to-Regulatory (T2R) macro-process to adjust the law to the needs that result from the evolution of technology or the relationship between intelligent systems and society. It allows for the implementation of a regulatory impact assessment.
- A Governance Committee to rule on the reports related to the impact of both R2T (*ex-ante* robot) and T2R (*ex-post* robot) processes.

- A data repository shared by R2T and T2R in order to gather data about whether each AI technology (planned or in use) is in compliance with the law.

Among the main benefits of this hybrid AI Governance Model, it is worth highlighting the integration of top-down and bottom-up regulatory actions in an incremental strategy, thus minimising the risk posed when regulating a new, constantly changing object.

The proposed Interactive Regulatory Governance Model helps to raise awareness regarding the lack of a continuous resource to connect both worlds—technology and legislation—while being iteratively developed and improved. Since the legislative branch is in charge of the legislation (in most democratic countries), it can be associated with the R-side of processes. When looking for the most ideal entity to act as the T-side of processes, the tasks of a regulatory agency can be identified.

Connecting both sides, R2T and T2R processes would be a strategy to establish a closer relationship between the legislative branch and the regulatory agency.

Ethics model for AI development and deployment (Schrader & Ghosh, 2018)

Founded upon philosophical principles and the dimensions associated with safeguarding human rights and well-being, the proposed ethical framework for AI development and deployment has been designed to implement core functions to represent ethical activities and the outcomes from both the philosophical and ethical perspectives.

The ethical perspectives are split into six categories: Rights (deontological ethics); Damages and Goods (teleological ethics); Virtue (aretaic ethics); Community (community ethics); Dialogue (communication ethics); and Flourishing (flourishing ethics).

The recommended core functions to be considered when developing AI systems are:

- Identifying ethical issues of AI—fairness, transparency, equity, goodness, beneficence, social utility, happiness, and protection of humans.
- Raising human awareness of AI—a clear understanding of how AI systems work within each product and how the industry develops algorithms.
- Collaborating with AI—dialogical interaction, listening, and understanding between humans and AI.
- Accountability of AI—guaranteeing the ethical compliance of AI systems and their designers.
- Integrity of AI—maintaining the AI system limited to the purpose for which the technology was intended.

A matrix combining the five core functions with the six perspectives has been built as a guideline to be followed during the AI project. As a proactive action in the design, development, and use of products and services that utilise AI, the model seeks to reflect the nature of social changes demanded by a new ethical thought.

Although they do not associate the framework with any specific organisation or institution, the authors' contribution can be applied by a regulatory agency when auditing the industry, as well as in its internal processes, to better understand the impacts technology has on the stakeholders.

Competency-based AI regulation model (Scherer, 2016)

Considering the competencies, strengths, and weaknesses of each state power, the proposal of an AI Regulatory Model (AIDA—Artificial Intelligence Development Act) is based on the distribution of responsibilities without losing sight of the mission goals. The model acknowledges the regulatory role of the executive, legislative, and judicial powers as agents in the regulatory process.

In the proposed model, the legislative branch would provide a statute placing a regulatory agency in charge of certifying AI products and services with regard to user and social safety. In general, legislators have limited knowledge of AI systems, their only support being a few committee meetings with experts. In order to solve this problem, legislators would delegate the responsibility for policy-making to the regulatory agency.

Supported by groups of researchers, the regulatory agency would comprise two main areas: policy-making and certification. Such an agency would be expected to be more agile and competent to monitor the evolution of technology, identify risks in the intelligent learning process and use of AI, issue technical recommendations, and verify that the technology is being applied for its intended purposes. A certificate would be given to designers, manufacturers, and service providers after being approved through the agency's processes. Pre-certification rules would also be made public to the industry and service providers. In case of an accident with certified products, the agency would publish a report to society, explaining the circumstances behind its occurrence and which certification rules/processes would therefore be modified.

Due to their *ex-post* nature, courts would judge cases considering whether or not a certification exists. Courts would judge companies for any losses and damages caused, considering the situation in which those organisations find themselves when it comes to certification. If a company's products or services cause any damages, if certified, the company would be judged based on more lenient

rules, whereas uncertified companies would be subjected to more rigid norms.

The proposed model takes into consideration the natural attributions of each entity within the government. Agility is required for the actions performed by the regulatory agencies, which would give them a prominent role in the regulation process. This is key to enable the evolution of technology while the legislation takes its time to mature.

Regulation model sustained by society (Rahwan, 2017)

Inspired by the Social Contract Theory (Rousseau, 2016), the Regulatory Model Sustained by Society adjusts the "human-in-the-loop (HITL)" to the "society-in-the-loop (SITL)" model.

The use of HITL thinking in AI has been largely applied to help an algorithm learn from humans' contributions. The agility and effectiveness of a HITL interactive learning machine stem from user feedback, thus enriching the knowledge that gets generated.

From a regulation perspective, the author argues that it is not sufficient to only adjust HITL to use a human to monitor an AI system and correct it in case of misbehaviour. By doing so, the regulation would rely on the judgment of an individual or group of individuals that subject the whole process to a narrow analysis. If we want to deal with a system that has an impact on the values of an entire society, that society must be included in the analysis, giving it a broader approach. It would not only avoid biased judgments, but also balance the competing interests of different stakeholders.

It is suggested that SITL be used in a process characterised by human-based government and citizen channels. On one side, the government's AI products and services would be run and, on the other side, citizens would evaluate those smart systems based on their own values. This would allow the government to understand how social behaviour and values change. Therefore, society-in-the-loop would become a governance tool for society to control and proactively identify those elements. Conflicts among safety-, privacy-, and justice-related concepts would benefit from this model. This relationship can be summed up as: society-in-the-loop = human-in-the-loop + social contract. The model also recommends auditing mechanisms to tackle the possibility of fake data manipulated by social groups at the learning stage, as well as results that would affect regulations.

For the purpose of using the proposed model as part of a broader AI governance model, both society and academia can be considered in terms of society's role when answering an agency's inquiry regarding the ethical behaviour of AI systems.

Principles of robotics (Boden et al., 2017)

After pinpointing the responsibilities of all agents involved in robotics, five principles were established in a guideline for robot designers, manufacturers, and users. The main goal of the rules is to emphasise that robots are tools, whereas humans are the actual responsible agents. The proposed rules are:

a. Robots should not be designed as weapons, except in the interests of national security.
b. Robots should be designed and operated to comply with existing laws, including those dealing with privacy.
c. Robots should be designed to be safe and secure.
d. Robots should not be used to exploit vulnerable users by pretending to feel emotions.
e. It should be possible to find out who is responsible for any particular robot.

Aiming to encourage responsibility within the robot-related research and the industrial community, seven messages have been created to highlight the responsible innovation spirit needed to abide by the rules.

The opportunity to use this proposal in audits performed by regulatory agencies can be identified, and that need must be reflected in the legislation to be adapted or created.

Agile AI governance (Wallach & Marchant, 2018)

Aware of the concerns regarding AI impacts exceeding the regulatory scope, capabilities, and jurisdiction of an agency or nation, the authors propose a model to address this governance challenge.

The model predicts actions performed by a Governance Coordinating Committee at the national level and a Global Governance Coordinating Committee. The main goal is a soft-law strategy that mitigates risks while the legislation is being drawn up. The soft governance part involves industry standards, social codes, labs, certification practices, procedures, and programmes. The hard governance part concentrates on laws, regulations, and regulatory groups.

A national committee would coordinate the efforts of a governance process encompassing stakeholders to produce recommendations, reports, and roadmaps, while monitoring those actions at the same time. This national forum would also be a perfect structure to enforce soft governance mechanisms as a necessary complement to the hard ones.

On the international level, a global committee would not only coordinate agreements among countries, but also establish a common understanding of which international standards should be used as a soft governance strategy. The international approach is also advocated to bring some balance to the several countries that are not yet participating

in the AI regulation dynamics, considering that the current situation makes them more vulnerable.

The proposed model takes a relationship network into account to address AI in a way that bolsters the formulation of actual standards while the legislation matures. The agile meaning of this governance is its incremental approach, which allows for continuous inputs. This would be an alternative to the problem posed by the temporal mismatch between formal regulatory actions and the production and commercialisation of deep machine learning-based products and services around the world. The success of this proposal depends on the amount of effort put into it by the market, academia, government, insurance companies, and organised civil society.

Sustainable AI development (Djeffal, 2018)

Considering the closer connection between sustainable development and governance, the author highlights that governance mechanisms are built to be continuously improved. The proposal concerns the entire lifecycle of an AI-based solution as the main foundation of a Sustainable AI Development (SAID) framework.

Analysed under the lens of a governance structure, SAID is stratified into the following layers: Technological, Social, and Governance.

At the base, the technology layer is in charge of specific applications involving architecture, data, and algorithm design.

Focusing on the impacts systems have on society, the social layer deals with the process of inserting technology into real life. It encompasses an analysis of the potential consequences of using AI in the social sphere.

Highlighting the importance of a broad treatment, the governance layer looks at the way algorithms influence both national and international decisions.

SAID gathers the different approaches examined in the various frameworks and somehow materialises the perception that, in order to be effective, AI regulation demands actions by IT and Social Sciences (Law, Business Administration, Philosophy, and Psychology) professionals alike. It also reminds us that, due to the topic's complexity, an AI governance model must include different process tiers.

Ethical framework for automation using robotics (Wright & Schultz, 2018)

Concerned with the integration between several stakeholders and automation using AI, this framework integrates the Stakeholders Theory with the Social Contract Theory in an attempt to find ethical grounds for developing, providing, and utilising AI.

The proposal considers as stakeholders: workers, the market, governments, the economy, and society in general. The impacts on the job market, from an ethical perspective, and the relationships among those stakeholders are highly emphasised.

The framework is based on a set of steps ranging from the identification of stakeholders, analysis of the social contracts among them, an assessment of how stakeholders are impacted, and lastly, actions aimed at mitigating the risk of terminating or breaching work contracts. An important target to be reached is increasing the benefits for stakeholders.

It is worth noting that this proposal considers as stakeholders those workers whose jobs or occupations will be modified with the introduction of AI into products and services. Due to the complexity of interests among stakeholders and all the labour concerns, the framework fits in the government policy-making process. The impact of such public policies on the country's economy may result in the need for laws, which means the legislative branch must be included as a stakeholder.

Intelligent model to regulate learning algorithms (Buiten, 2019)

Focused on a strategy to fight intelligent services that contain biases, this model postulates that an algorithm should assess the essential elements of a machine learning process (data, testing algorithms, and decision models). The proposal is founded on the thesis that the transparency of a code is insufficient to guarantee an unbiased solution and admits that it is still possible to find biases, even when learning from vast amounts of data.

In the data domain, all data samples are assumed to include some built-in biases that need to be considered. The data must be checked to ensure their validity, reliability, and proper data dependency.

Regarding the testing algorithms, the model recommends using a variety of algorithms and comparing their performance. However, that must be done only after discovering the quality of the available data.

The decision-making process is seen as a delicate phase in which developers must be aware of the correlations between variables, because hidden relationships may obscure a biased orientation. It also acknowledges the difficulty of identifying those problems automatically as algorithms grow in complexity.

Universal declaration of human rights as a framework (Donahoe & Metzger, 2019)

This model is founded on the argument that the several different frameworks related to each specific area of ethics are insufficient to regulate AI on an international scale,

both in the private sector and within the government. Due to that gap, the Universal Declaration of Human Rights (Kunz, 1949) has been considered a mature approach that different cultures have been adopting for decades. Modern adjustments were made by the UN Human Rights Council in 2011, published as the UN Guiding Principles on Business and Human Rights (United Nations, 2011), which highlight the roles and responsibilities of private-sector businesses in the protection of human rights.

Under the human rights framework, governments have the duty to protect citizens from violations and infringements of their rights by other governments and non-State actors, including the private sector. Donahoe and Metzger's proposal deals with the centrality of the human person as the focal point of governance and society. It seeks to address the potential impacts of AI, such as:

- The right to equal protection and non-discrimination—avoiding biases in the data and ensuring fairness in machine-based decisions.
- The right to life and personal security—concerning autonomous weapons that move beyond human control.
- The right to an effective remedy for violations and infringements of rights—transparency, fairness, and accountability in cases where AI systems impact people's rights.
- The right to privacy—addressing the loss of privacy in data-driven societies and the need to protect personally identifiable data.
- The rights to work and to enjoy an adequate standard of living—guiding governance decisions around the displacement of human workers by AI.

Software requirement model for the ethical assessment of robots (Millar, 2016)

Considering ethics as a social enterprise, the proposal puts forth a set of general specifications to be considered in a system aimed at assessing robots during their construction. To that effect, five major rules have been built:

- Balancing designer and user requirements, considering the potential damages.
- Utilising a user-centred ethical evaluation tool for AI systems, which must use design methodologies that are able to identify the impacts on human values in use contexts.
- Including the psychology of user-robot relationship variables in the ethical evaluation tool to identify variables such as the user's emotional state.
- Compliance with the Human-Robotics Interaction Code of Ethics (Riek & Howard, 2014).
- Designers' understanding of both acceptable and unacceptable design features, which could be implemented by including ethicists in design teams.

It seems the proposal may be utilised by the industry and regulatory agencies alike. In both cases, it could be the first red flag signalling the need for a red button in robot projects (Arnold & Scheutz, 2018).

Ethical judgement model for codes (Bonnemains et al., 2018)

Considering that (a) an ethical framework allows us to deal with situations involving ethical dilemmas, (b) one framework alone is not efficient enough to compute an ethical decision, and (c) tackling ethical decisions is better than avoiding them, the author proposes a formal logical model that can be implemented by an agent facing an ethical dilemma, with the ability to both make decisions and explain those decisions. It assumes that formal expression analyses are especially useful to identify the subjectivity of a decision.

Different judgements on possible decisions have been studied according to three ethical frameworks: consequentialist ethics, deontological ethics, and the Doctrine of Double Effect. In the path toward a refined and final framework, various ethical dilemmas have been formalised in judgment functions that return three possible results: acceptable ($\top$), unacceptable ($\bot$), or undetermined (?). The concepts of 'decision', 'event', and 'effect' were taken into account when building the model's functionalities.

Those analyses can be appreciated when we judge someone or something based on particular moral theories.

Asilomar AI principles (Future of Life Institute, 2019b)

The governance model proposed by the Asilomar Conference resulted in 23 AI Principles undersigned by thousands of experts (Kozuka, 2019). Grouped under "Research Issues", "Ethics and Values", and "Longer-Term Issues", those principles encompass the lifecycle of an AI-embedded product or service—from motivation and funding to the assessment of benefits and judgement criteria concerning its impacts.

In the Research Issues dimension, the recommendations are to: research goals and funding, establish a connection between researchers and policymakers, research the culture of cooperation, and promote synergy to avoid corner-cutting when devising safety standards.

In the Ethics and Values dimensions, the orientations are to: maintain AI systems secure during their entire lifecycle, make them transparent in case of failure as well as

in judgment results, consider designers and builders of advanced AI systems as stakeholders in the responsibility chain, align AI systems' values with their users', design AI systems to be compatible with human rights and cultural diversity, preserve personal privacy, share their benefits as much as possible, and make it possible for a human to take control of AI systems, if so desired.

Finally, in the Longer-Term Issues sphere, the principles are to: be cautious when making decisions without a consensus, build a mitigation plan to deal with the risks, plan a recursive type of self-improvement, and develop AI systems based only on widely shared ethical ideals.

European ethics guidelines for trustworthy AI (AI HLEG 2019b)

With the goal of creating guidelines to orient a new AI governance, a team of experts entitled High-Level Expert Group on Artificial Intelligence has drawn up the Ethics Guidelines for a Trustworthy AI for the European Commission based on a structure supported by values that should be considered throughout the system's lifecycle: lawful, ethical, and robust AI.

Based on the European Union Charter of Fundamental Rights (EU Parliament, 2012), the model establishes trustworthy AI as a key element for a governance framework (Kozuka, 2019) has been built using a three-tier structure.

The highest tier addresses ethical principles based on fundamental human rights: respect for human autonomy, prevention of damages, fairness, and explicability. To ensure fairness in a society with different interests and objectives, it defends an explicable decision-making process. It should consider traceability, auditability, and transparent communications regarding system capabilities. It also recommends that particular attention be paid to vulnerable groups and situations characterised by asymmetries of power or information (employers and workers, or businesses and consumers).

The second tier includes the key requirements necessary for implementing an AI-based system or service throughout its lifecycle: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; social and environmental wellbeing; and accountability. All requirements are connected to one another through a full-mesh relationship where each one of them has the same weight.

Special attention is suggested to the oversight as part of a governance mechanism that could use human-in-the-loop (HITL), human-on-the-loop (HOTL), or human-in-command (HIC). A strong connection can also be found between "privacy and data governance" and "diversity, non-discrimination, and fairness", due to the need for mechanisms to avoid inadvertent historical biases, incompleteness, and inadequate data governance models. Regarding the accountability concerns, a recommendation is given to carry out an impact assessment prior to and during the development.

Defending a trustworthy AI implementation throughout the lifecycle of an AI system, the model demands a process-oriented approach that encompasses both technical and non-technical methods when implementing the requirements. Within the non-technical approach, one can find legislation and corporate guidelines encompassing codes of conduct, policies, performance indicators, and agreed-upon standards. Those standards consider AI users, consumers, organisations, research institutions, and governments as stakeholders. They also include a certification granted to organisations that produce transparent, accountable, and fair AI systems in accordance with the established standards. The entity in charge of the certification could play an important role in the communications with "industry and/or public oversight groups, sharing best practices, discussing dilemmas, or reporting emerging issues of ethical concerns."

For the base tier, a list of recommendations directed at the operationalisation of the key requirements in the upper tier for each specific system has been formulated.

Ethically aligned design (IEEE 2019)

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems has proposed five general principles for AI systems and a guideline recommending actions to establish ethical and social implementations for intelligent and autonomous systems that prioritise human wellbeing.

According to that model, the ethical design, development, and implementation of AI systems should consider the following principles: human rights, wellbeing, accountability, transparency, and awareness of misuse.

Focusing on personal data rights, the wellbeing promoted by the effects on the economy, the legal frameworks for accountability and transparency, and the education and awareness policies, recommendations were made to a wide set of stakeholders.

To governments: the governance framework should include standards and regulatory agencies, provide society with ethics education and security awareness regarding the potential risks, improve digital literacy, use multiple metrics as wellbeing indicators, and implement a wellbeing impact assessment.

To industries: programmatic levels of accountability should be provided to address culpability in legal matters, transparency by design, intelligibility of a system's operation and decisions, damage mitigation strategies, assessment starting at the design phase, understanding how each jurisdiction would treat the damage caused by a given AI system.

To the legislative sphere: responsibility, culpability, liability, and accountability issues should be classified.

General recommendations: certification of AI systems, identification and prioritisation of standards for each category of AI systems, continuously updating the standards, metrics utilised to assess AI systems, agreements on moral decisions, evaluation by third parties, applying the classical methodologies of deontological and teleological ethics to machine learning, adherence to the code of conduct by the AI production team, and bridging the language gap between technologists, philosophers, and policymakers.

At the international level: establishing a global multi-stakeholder dialogue to determine the best practices, facilitating AI research and development in developing nations, and using indicators to assess AI-related technological interventions in those countries.

Government and industries: identifying the types of decisions and operations that should never be delegated to AI systems.

A special guideline for implementing an ethical culture in organisations (IEEE, 2020) has also been built, encompassing a strategy to assess the level of each dimension to be developed (lagging, basic, advanced, and leading).

Avoiding biases and discrimination (Lin et al., 2020)

In order to amplify the effectiveness of bias-reduction intervention procedures in cases of implicit biases, the framework explores an innovative AI-assisted intervention based on a bidimensional approach.

In the first dimension, the different types of information AI provide to users are captured: the current state of affairs (descriptive information), the likelihood of future states (predictive information), and the expected utility of an action (prescriptive information). It considers that all interventions are prescriptive, and the knowledge-based systems (KBS) will decide to intervene depending on how they simulate the results.

In the second dimension, an AI system can intervene in different phases of the decision-making process (input-based interventions, output-based interventions, and cognition-based interventions) as part of an interactive process.

It is a case of regulation by software, which could be used by the industry and service providers as part of their internal process.

Standardisation exchange model (Lewis et al., 2020)

Considering the importance of standardisation in a regulation strategy, the model proposes a process among functional entities in the AI value chain through which information related to standards is exchanged among them.

Classified by their functional roles, the actors—data providers, AI system creator, AI system operator, AI user, oversight authority, and associate stakeholder—change standards focusing on a trustworthy AI.

The benefits of each exchange are presented, as well as the potential topics for new standardisations. Most of them concern issues to be considered in an AI product certification process.

Although the focus is on the industry, the model considers the importance of the government in the whole process and the need for an international community to discuss the standards.

Algorithmic impact assessment (Canadian Government 2020)

Aiming to help public and private-sector companies assess and mitigate the impacts of deploying an automated decision-making system, the Canadian Government has developed the Algorithmic Impact Assessment (AIA) based on the Government Directive on Automated Decision-Making. The AIA questionnaire considers the reasons for using AI on decision-making processes, the capabilities encompassed by the system, algorithm transparency and explainability, system category (health, social assistance, economic, etc.), development and training process, system and data architecture, stakeholders, and risk mitigation measures.

The impact assessment addresses the four levels according to how the decisions impact the rights, health, or well-being, the economic interests of individuals or communities, and the ongoing sustainability of an ecosystem. Thus, levels I, II, III, and IV are each related to a certain impact, namely, reversible brief, reversible in the short term, difficult to reverse, and irreversible.

The Directive on Automated Decision-Making was designed by the Canadian Government to make its administrative decisions compatible with core administrative law principles, such as transparency, accountability, legality, and procedural fairness.

The requirements considered by the Directive on Automated Decision-Making are distributed between two pillars: transparency and quality assurance. Among the transparency requirements, it establishes that:

- Notice on relevant websites must be issued before decisions are made,
- Meaningful explanations must be provided to affected individuals regarding the decisions made,
- The Government of Canada has the right to access all components of the system.

Among the quality assurance requirements, there are rules to ensure testing and monitoring outcomes, data quality, peer review, employee training, contingency, security, compliance with the law, and human intervention.

AI governance by human rights-centred design, deliberation and oversight (Yeung et al., 2019)

Considering international human rights-based standards as the most promising governance framework to deal with ethical standards, Yeung et al. (2019) have proposed the Human Rights-Centred Design, Deliberation, and Oversight model to deal with AI-related ethical issues with legal support. Based on a global approach, the proposed model integrates a suit of technical, organisational, and evaluation tools and techniques involving many stakeholders.

The proposal presents norms based on human rights as the foundation for ethical standards with which AI systems must demonstrably comply:

a. Design and development that take stakeholders' opinions into account. In case an assessment has resulted in "high" or "very high" risks to human rights, a redesign should be pursued.
b. Formal assessment and testing to evaluate their compliance with human rights-based standards. It would occur regularly during the entire life cycle of a system's development—design, specification, prototyping, development, and implementation. A systematic and periodic post-implementation monitoring would be established, through which the AI system would be submitted for review by sending out the related documentation and reports to a public authority.
c. Independent oversight by an external, technically competent entity invested with legal investigation and sanction powers.
d. Auditability supported by traceability and by evidence that the AI system is operating as desired and that it was properly documented during its entire life cycle of development.

The authors highlight the need for laws and norms encompassing all steps covered by the model.

Good AI society (AI4People, 2018)

Focused on the establishment of a good AI society, the proposal joins ethical principles and specific recommendations to enable stakeholders to seize opportunities and avoid or minimise risks.

The model encompasses five ethical principles: Beneficence, Non-maleficence, Autonomy, Justice, and Explicability.

The recommendations are categorised as: assessment, development, incentivisation, and support.

- Assessing institutions on their capacity to reduce the mistakes made by AI systems.
- Considering existing legislation, using participatory mechanisms to align with social values, and assessing tasks/decision-making that should not be delegated to AI systems.
- Assessing current regulations to provide a legislative framework that could keep pace with technological developments.
- Developing a framework to enhance the explicability of AI systems.
- Developing legal procedures to permit the scrutiny of algorithmic decisions in court.
- Developing auditing mechanisms for AI systems to identify unwanted consequences.
- Developing a process to remedy or compensate for damage caused by AI.
- Developing agreed-upon metrics for the trustworthiness of AI products and services.
- Developing a new EU oversight agency responsible for the scientific evaluation and supervision of AI products and services.
- Developing a European observatory for AI.
- Developing legal instruments to prepare and adjust the work environment to the changes brought about by AI.
- Financially incentivising a socially preferable development and use of AI.
- Financially incentivising cross-disciplinary cooperation in the fields of technology, social issues, legal studies, and ethics.
- Incentivising a regular review of the legislation to foster socially positive innovation.
- Financially incentivising the use of lawfully special zones for empirical testing and development.
- Financially incentivising research on the public perception of AI.
- Supporting self-regulatory codes of conduct for data- and AI-related professionals.
- Supporting corporate boards of directors to take responsibility for the ethical implications of AI technologies in their organisations.

Framework approaches

An analysis of the approaches adopted by each of the 21 frameworks proposed in the sample resulted in Table 1.

The fact that ethical guidelines exist is not enough to have any effect on the software development industry. Thus, models that are strongly grounded on ethical principles require legal mechanisms to fulfill those recommendations (Hagendorff, 2019).

Frameworks that encompass the competencies of government institutions have also foreseen the existence of a regulatory agency, as well as the need for mechanisms to

Table 1 Comparative table of the approaches explored in the frameworks, compiled by author.

Approach	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	5.10	5.11	5.12	5.13	5.14	5.15	5.16	5.17	5.18	5.19	5.20	5.21
Institutional competences		■		■			■								■	■			■		■
International							■	■			■				■	■		■		■	
Hybrid (soft + hard)		■		■			■							■					■	■	■
Successive interactions		■												■							
Regulatory agency		■		■			■								■				■	■	■
Gradual improvement	■	■		■	■		■							■	■	■					
Ethical principles	■		■			■			■		■	■	■	■	■	■			■	■	■
Social contract					■				■												■
Job Market									■							■					■
Impacts on stakeholders		■	■			■	■		■		■			■	■	■		■	■	■	■
Governance		■	■	■			■	■							■	■			■	■	■
Process based		■							■	■					■		■	■	■	■	■
Technology as a regulator										■			■				■				

help the legislative branch speed up its law-making process, aiming for a safer and faster AI regulation.

Frameworks that take the social contract into account rank among the most open to society's participation in a co-production with the government. Those models consider citizens as outstanding stakeholders. Concerns over the impacts on the job market are also a way to assess the impact on stakeholders.

The main argument that proposes a gradual deployment of the regulation is a risk mitigation strategy, but it could also be combined with successive interactions between the legislative branch and the regulatory agency, thus enabling continuous improvement during the legislative procedure.

The interactive regulatory governance model, the agile governance, the ethics guideline for trustworthy AI, the ethically aligned design, the algorithmic impact assessment, the good AI society, and the AI governance by human rights-centred design, deliberation, and oversight proposals encompass a larger number of topics. The AI HLEG proposal highlights that a trustworthy AI must be lawful, ethical, and robust. The others explore the relationship among all parties involved in the regulation process and the attempt to find balance between more or less rigid or flexible mechanisms. It is worth noting that the agile governance proposal does not exclude conventional actions for a formal regulation—the interactive regulatory governance model and the competency-based regulatory model, both of which involve the legislative branch. Therefore, this configures a transitional situation in which consensual standards would be agreed upon and enforced, and the risks would be mitigated until legal mechanisms are made official, which is very similar to the concept of Dynamic Regulation, in which feedback serves as a basis for the maturity of the regulatory instrument (Kaal & Vermeulen, 2017).

When analysing several movements advocating the establishment of criteria for best using AI, studies identified an opportunity to develop a competition around a technological reform (Greene et al., 2019). Pondering over the need to find synergy among global AI regulation-oriented actions, a few proposals rely on a worldwide effort, which sometimes is described as an international committee, while other times just as a joint effort by governments and multinational companies.

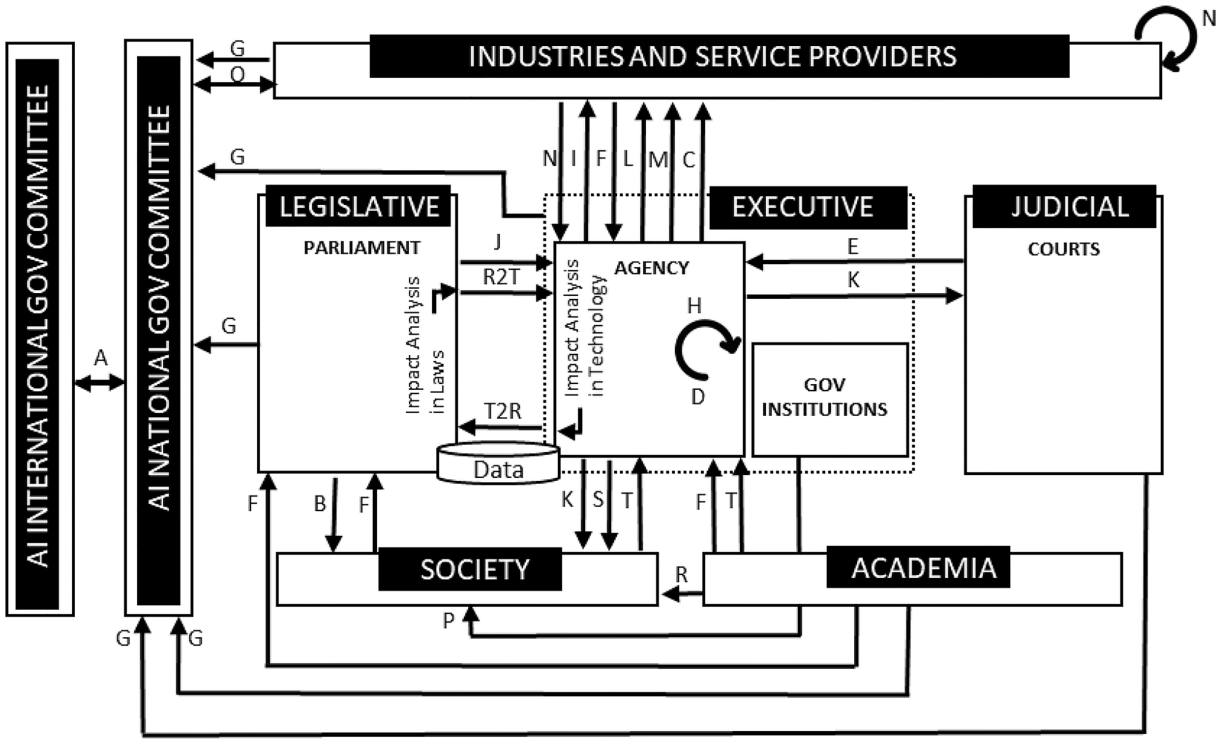
Despite the small number of existing software-based regulation models, similar models are likely to arise, since the increasing complexity of AI solutions results in more system rules (Lamo & Calo, 2018; Liu, 2017; Prakken, 2017; Verheij, 2016), which in turn means a higher likelihood of conflicts among those rules in combined systems (Bench-Capon & Modgil, 2017).

AI regulatory and governance framework

The supplementary nature of some of the models confirms the perception that the impacts of AI would demand a combination of design, laws, and education (Calo, 2011). When debating over the complexity of a framework to address such a multidisciplinary topic (Bonnemais et al., 2018) embedded into the political and social context (Leitner & Stiefmueller, 2019), an AI regulatory and governance framework—AIR—was built to include the main contributions from each model in the examined sample (Fig. 1).

Focused on reducing the gap between ethical principles and actions by each stakeholder and making the relationship among them in different dimensions of knowledge clearer, the AIR framework is based on a wide governance process.

Although the government's exclusive competencies are highlighted, aiming for more accuracy in its actions, the power of the State has been distributed among the legislative, the executive, and the judicial branches. This



A – International agreements
B – Laws and bills
C – Certification
D – Standardisation researching process
E – Results of judgments
F – Feedback and contributions
G – Participation in committees
H – Certification process
I – Certification rules
J –Agency creation satute
K – Certified products and services
L – Auditing process
M – Algorithm Impact Assessment Questionnaire
N – Industry standards
O – Risk management standards
P – Public policies
R – Risk analysis report
S – Report about incidents with certified products and services
T – Answer to consulting about system behaviour
R2T - Regulatory-to-Technology process
T2R - Technology-to-Regulatory process

Fig. 1 AIR framework

segmentation is used by many countries as a functional way to distribute power, according to which the legislative creates the laws, the executive enforces those laws, and the judicial is in charge of solving whatever conflicts arise to guarantee justice and law abidance (Maluf, 1995).

Apart from making laws, it is crucial to maintain the legislative branch open so that its bills (B) can be discussed with society, receiving constant feedback and contributions not only through e-participation systems, but also through a special channel established with scholars, who could also attend the legislative committee meetings (F).

The Parliament or Congress, as an instance of the legislative branch, would approve a statute (J) to create an AI regulatory agency as part of the Federal Government (executive branch). This could be a good moment to define AI, or at least to demand that the agency do it.

Upon its creation, the regulatory agency would establish a strong relationship with the Parliament as part of an ongoing process in which the legislative would survey the impact on the legislation and its evolution based on the knowledge obtained from the regulatory agency (T2R—Technology-To-Regulatory), much like the regulatory agency structures its internal work processes based on the legislation discussed and approved by the legislative (R2T—Regulatory-To-Technology).

The T2R is necessary, at least until each new category of AI systems has been deeply studied by the regulatory agency. Due to the complexity and specificity of AI services and products, laws could potentially be created for each

specific field. The natural evolution of the former would also cause the latter to evolve in the long term. A practical way to implement the T2R flow is through the regulatory agency frequently attending the legislative committee meetings to discuss AI regulation.

As a complement to T2R, the R2T flow would be started at least when a new version of a bill is discussed at the legislative committee meetings and when a new law is approved. R2T also feeds other internal processes of the regulatory agency in order to update them with the legislative understanding of what can be regulated by law, which can trigger three reactions: (a) alerts regarding the limitations that the bill/law brings to the ongoing projects of the industries and service providers; (b) opportunities to expand the standards by discussing them with the industries and service providers; and (c) updating certification and auditing processes with new compliance issues.

Among the regulatory agency's competencies, a couple of processes require speed and synergy: analysing how the legislation affects the technology process and standardising the research, certification, and auditing processes. In order to be effective, those processes must consider a huge number of variables involved in the entire lifecycle of an AI system: design, prototyping, development, testing, deployment, commercialisation, and use. The efficiency and knowledge of the regulatory agency are expected to possess depend on mechanisms that support those processes (Fig. 2): formal representation models for ethical dilemmas, impact on stakeholders' assessment according to ethical principles, data governance assessment, development process assessment, systems to identify biased machine learning, and assessment of risk mitigation measures.

Being responsible for a closer interaction between the Parliament and the regulatory agency due to the R2T flow, the analyses of how the legislation affects the technology process must be corroborated with information structures that are able to represent the law based on a technical mindset. Those involved must be skilled in both areas of knowledge. The quality and efficiency of this synchronicity of mixed mindsets are strengthened by means of a data repository shared by the Parliament and the regulatory agency. Examples of such data would include: issues related to whether AI projects/products comply with the law, AI project/product impact assessment, legislation/regulation collected overtime across AI projects/products, ethical committee decisions upon approval requests, AI project/product and regulatory assessments under both private and public guidelines.

In order to offer controlled autonomy to the AI industry, technical standards must be established while bills are being discussed. An agile interaction between the "industries and service providers" and the regulatory agency is supported by the standardisation of the research process (D). Despite being a process inside the regulatory agency, standardising research actions entails an in-depth study of ethical and safe mechanisms to make the new projects seen in the AI market feasible. A very technically skilled staff must be allocated to that task, which requires robust laboratories.

As a strategy to motivate the AI market to follow the best practices, standards, and laws (when they exist), the regulatory agency would certify products and services using a certification process (H). Companies that submit their products to the regulatory agency, after a successful appraisal, would receive a certificate (C) within their field of action (transport, healthcare, entertainment, education, military,

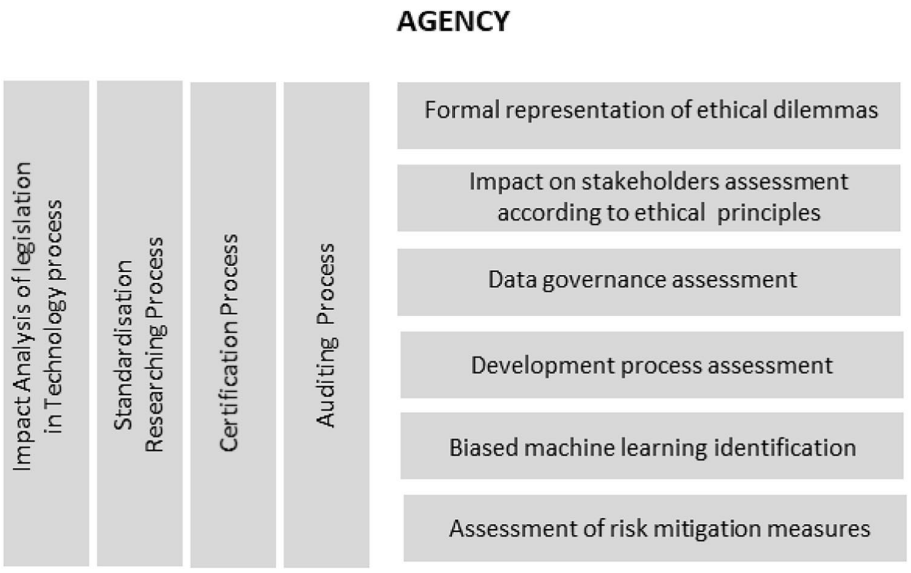


Fig. 2 Regulatory agency in the AIR framework

etc.). The strictness and nature of the assessment process could be different for each of those fields. It is also a way to communicate to society where people can place their trust when buying or using an AI system.

Through a quick process, the industries and service providers would need to receive the regulatory agency's certification rules stated as clearly as possible (I), while providing feedback (F) on the conditions that preclude the development process required by the regulatory agency from moving forward. Accountability requirements would be assured if those lists would show not only new products and services that have been certified, but also those that have lost their certification.

The issuance of certificates could be a strategy to be applied before laws are passed, since they already inform society, in a transparent fashion, about the safety levels and risks of the products and services it consumes. Advertising campaigns by the government and certified companies would also strengthen that strategy.

A robust strategy to avoid fake certifications would be desirable, such as a blockchain mechanism implemented by the agency containing the updated certification list for a given country. Aiming to increase citizen trust in AI certification, certificates could be issued using a digital signature in the system's code, setting an attribute associated with that specific code version. In case someone wants to know if the version of a commercialised AI system is updated, all they need to do is compare the digital signature with the one that is available on the regulatory agency's website.

The regulatory agency could make an "algorithm impact assessment questionnaire" (M) available to the industries and government institutions in order to offer a simulation tool through which they could know, in advance, their level of compliance. It would also fit as a preparation stage for a certification submission.

And finally, an auditing process (L) would be supported by the regulatory agency to check companies demanding certification and certified companies that need to update their certification, as well as to verify issues demanded by courts in case of sentences related to damages supposedly caused by AI systems. This audit would take place in five dimensions: impact on stakeholders based on ethical principles, data governance, development process models, identification of biased machine learning, and risk mitigation measures. The auditing process should be part of regular monitoring through which not only internal changes in companies, but also future problems coming from new arrangements in society could be identified.

Any failures or damages noticed in a certified AI product or service must trigger an internal audit process to identify whether there were problems or limitations in other agency processes that could be a reminder of internal improvement. In a broader, more transparent fashion, the agency should publish the audit results and the next steps (S).

The regulatory agency's processes are interconnected through a knowledge stemming from the mechanisms shown in Fig. 2, which should be handled as much as possible by a skilled multidisciplinary team, since the ethical and technological dimensions are mixed.

Mechanisms for formally representing ethical dilemmas are important to create a transparent communication channel between ethicists and technical profiles. It could also help distinguish between the part of the decision-making algorithm that is related to a dilemma and the rest of the code in relation to which there is a consensus regarding the best decision. This representation model is expected to be continually improving as society changes and new dilemmas are identified. This analysis is interconnected with the impact on the stakeholders' assessment according to ethical principles.

As each company has its own system development process, the regulatory agency must have a process to guarantee a broad system development process assessment, probably by attempting to measure the sample against the best practices and the risks related to each step that does not follow them.

Since a biased machine learning can result from problems with data collection, testing algorithms, or decision models, the regulatory agency must consider all those phases in its development process assessment models. A data governance assessment is an important analysis that is connected to the system development process as well as to the biased machine learning identification process.

The results of the regulatory agency's analysis materialise the total sum of all risks identified in an evaluated AI product or service for which there should be a risk mitigation plan.

As the regulatory agency is a natural actor to create and communicate the best practices to the industries and service providers, the agency must be aware of all projects and trends in the AI market, otherwise companies will not adopt those practices. An alternative to mitigate that risk is to strengthen the dialogue with the industries and service providers on the purpose of contributing to industry standards (N), thus allowing technology to improve its development while the legislation is still under debate, or in case it is not necessary. On the industries' and service providers' side, in order to increase the probability of a successful investment, a gradual strategy supported by a governance model should be behind the implementation of those good practices. Industry standards (N) must incorporate all parameters that are needed for communication among the "industries and service providers" along the entire value chain of an AI system.

As happens in Parliament, an open practice by the regulatory agency is likewise desirable, receiving feedback from academia (F). That feedback and those contributions, among

other information, could be how society perceives ethical behaviours. A partnership between academia and the regulatory agency, combining scholars and researchers in the agency's staff, could be a sustainable alternative for maintaining a highly skilled team of professionals dealing with many processes simultaneously.

At an advanced level of an AI governance model, a "society-on-the loop" mechanism could be structured to collect the evaluation of a certain category of AI systems based on their behaviour using an ethical approach. Both civil society and academia could accomplish this. The answers (T) would feed the regulatory agency in the form of a survey to identify potential opportunities for improvement in its internal processes.

Regardless of the existence of a "society-on-the loop" mechanism, academia is always a good, reliable source of risk analysis reports (R) to be published periodically.

Law enforcement by courts would also undergo a continuous learning process with regard to interpretations based on the legislation in effect, as well as on new laws. In countries where the certification is incorporated into laws, decisions on cases involving uncertified companies would be treated differently from those involving certified companies. Thus, society and the courts would need to have up-to-date information about each company's certified products and services (K). Considering a continuous learning process, the regulatory agency would receive the judgment decisions of all cases involving AI systems (E), which would then be stored in the data repository shared with the Parliament. The decisions on cases may indicate types of AI technology use that the regulatory agency has not researched yet, and they may also indicate the need for changes in the legislation. A significant challenge would be to identify when an incident is avoidable or not. In those situations, experts must be involved in the investigation to find out the purpose of supporting the courts.

In order to balance the equation that rules the job market on the path to a digital economy, the government may create public policies (P) to make it feasible to implement in a timely manner the changes required in employer and student skills. Public policies might also be necessary to maintain an advertising campaign to inform people about the importance of certification and standards for AI products and services, helping them to identify when there is a potential case of an AI-embedded system.

As usual, public policies are a long-term strategy that may require actions by different government institutions, but there are many alternatives for implementing them, depending on the country. The regulatory agency may also provide government institutions with information about where and how those changes are needed. In some cases, by means of the T2R flow, the agency may notify the Parliament that a law is lacking that better regulates public policies.

On a national level, discussions to facilitate priority actions and the recognition of industry standards would be enabled through an AI Governance Committee, bringing together the public and private sectors (G). The synergy of efforts for the benefit of all stakeholders must be established, since many variables are considered. Beyond the regulatory agency, other government institutions would probably participate in this national committee, due to the wide impact its decisions could have. For instance, building human capacity and preparing the labour market transformation is a decision that might require a strategy that impacts many ministries and state governments. Adjustments to the current legislation related to many different subjects should probably be made to support the whole transition.

We should not forget the committee's governance approach, which requires working with indicators, i.e. data produced by its stakeholders. Therefore, a national AI governance committee would require at least collection, storage, and analysis processes within other institutions and businesses.

The agreed-upon standards (N) make it possible to move forward in some technological dimensions, while the Parliament discusses adjustments to the legislation when necessary. The risk management criteria (O) related to the use of those standards would be negotiated between the national committee and the industries and service providers, since each standard could impact a long productive chain.

The plethora of components in AI services and products of global reach imposes actions that would be agreed upon in an International Governance Committee comprising representatives from each country's committee (A). On many occasions, transparency in production processes is only feasible through complex international agreements, because corporate trade policies must adapt to different countries. A global strategy could be established to facilitate the production and delivery of standards, as well as the dissemination of best practices in undeveloped countries, since without that help the gap between them and the countries in which an AI governance has been established would increase hugely, putting them in a fragile position. In that regard, one should keep in mind that international standards are not limited to technological issues. Further, those standards also incorporate ethical principles, despite any cultural differences. The Universal Declaration of Human Rights could be a global base to engage governments to face the challenge of dealing with differences among national legislations.

The expert skills and engagement power of self-regulated organisations are a rich contribution to the international AI governance committee.

A possible adjustment entails the segmentation of tasks in charge of the regulatory agency, sharing them with or

transferring them to other government institutions. For instance, the audit process could also be implemented by different government institutions in charge of auditing cases of discrimination using personal data, or investigations related to the development of autonomous weapons in that country. Hence, it is important to highlight that laws such as the EU GDPR (2016) only affect personal data. Nonetheless, AI discrimination risks have a wider reach than personal data.

Sharing the standardisation process with specialised private-sector organisations could also be an alternative. In that case, the connection between the standardisation process and the other regulatory agency processes should be maintained.

Despite being represented as a unique institution, the regulatory agency could be materialised as a group of agencies distributed across the country. To that effect, partnerships among countries could also allow for the creation of a set of agencies sharing resources, processes, and knowledge. In both cases, agencies could specialise in different categories of AI products and services. Although the certification issued for a specific category is independent of the certificate issued for another category, a communication process among the agencies is needed to increase the knowledge of how each AI product/service behaves and evolves over time.

Another adjustment to how the AIR framework is interpreted relates to what can be classified as "industries and service providers". Private-sector companies are considered first. However, since any organisation that develops AI systems or offers services based on AI systems would fall in that category, public organisations may also be included.

Conclusion

The need and urgency to regulate Artificial Intelligence seem indisputable. The complexity of the topic is also evident, whether due to the advanced nature of technology or because its impacts structurally affect social standards. This combination materialises the perception of a problem that is yet to be completely defined.

A study of the literature through a sample comprising 109 documents (articles, laws, and government strategies) revealed significant efforts to identify and scale the risks and ethical dilemmas related to AI, as well as to seek a model for regulating AI based on different methodologies.

The heterogeneous nature of the professional profiles involved in the debate evinces the complexity and maturity with which the topic is being studied. Such an in-depth approach, on the one hand, may have caused certain delays in research, but on the other, it has prevented inappropriate regulatory solutions from being made official.

We had also seen the birth of a reshaped perception of the legislation, as had occurred with disruptive innovations in the past, when legislative efforts focused on adapting laws to the new paradigms brought about by electricity, telephone, and computers. Since this is a more difficult challenge, AI lawmakers will consider that we are still starting to discover the applications of smart algorithms. Therefore, a balance must be kept between a rigid damage prevention and technological development strategy (Gurkaynak et al., 2016).

Despite all efforts being directed to AI regulation and governance, there is still an expressive gap between ethical principles and a functional model that is able to encompass all areas of knowledge that are necessary to deal with the required complexity. The 21 proposed models found in the sample are based on supplementary approaches and are therefore insufficient when analysed separately. Due to the heterogeneous nature of those skills and interests, an ideal model should harmonise interests, offering benefits to all stakeholders during the entire lifecycle of an AI product or service.

The consolidation and process orientation approach proposed by the AIR framework (Fig. 1) seems to be the most adequate strategy for the deployment of an AI governance, given the existence of several agents and the laterality of the topic, which intertwines different areas of knowledge. The expanded view of the presented AIR framework will enable all agents involved to identify their role in the governance process, while establishing a roadmap for a gradual and uninterrupted deployment.

It also contributes to the creation of a new reward and punishment model to balance out this new reality (Bryson, 2018; Waser, 2015), taking into account the world as it will be (Lin et al., 2011).

On the path to improve each component of the AIR framework, more than bringing them closer together, there needs to be a synchronisation of stakeholders towards a sustainable regulation. Along that journey, an alliance between scholars and the government's three agents (the executive, legislative, and judicial branches) is crucial for the macroprocess of regulation.

The countries leading the debate are probably ready to coordinate the partnerships and agreements among institutions that are necessary for a comprehensive and effective governance, as well as to initiate a regulation process. Nonetheless, the launch of AI-embedded products in countries that have advanced regulation models, in and of itself, does not guarantee the same safety levels for countries that are still unripe in this regard.

Much is yet to happen in the formulation of solutions using real-case scenarios to enable an empirical analysis and studies of the evolution of the models presented in the examined sample. To that effect, the AIR framework can make it tangible and feasible to synchronise all the stakeholders' efforts to achieve an effective result, thus culminating in the creation of a reference model of AI

governance in which maturity levels would be established that could be monitored by international bodies in a collaborative action. The way we and future generations will live our lives depends on that cooperation.

References

Aayog, N. (2018). National Strategy for Artificial Intelligence: #AI for All (Discussion Paper) https://www.niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf. Accessed 30 July 2020.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access Review, 6*, 52138–52160

AI HLEG - High-Level Expert Group on Artificial Intelligence. (2019a). A definition of AI: Main capabilities and disciplines. Definition developed for the purpose of the AI HLEG's deliverables.

AI HLEG - High-Level Expert Group on Artificial. (2019b). Ethics guidelines for trustworthy AI. High-level expert group on artificial intelligence for the European Commission.

AI4People. (2018). Ethical framework for a good society: opportunities, risks, principles, and recommendations. *Atomium – European Institute for Science, Media and Democracy.* http://www.eismd.eu/wp-content/uploads/2019/02/Ethical-Framework-for-a-Good-AI-Society.pdf. Accessed 21 June 2019.

Amigoni, F., & Schiaffonati, V. (2018). Ethics for robots as experimental technologies. *IEEE Robotics & Automation Magazine, 25*, 30–36

Arkin, R. C. (2011). Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, 121–128.

Arnold, T., & Scheutz, M. (2018). The big red button is too late: an alternative model for the ethical evaluation of AI systems. *Ethics and Information Technology, 20*, 59–69

Beltran, N. (2020). Artificial intelligence in Lethal Autonomous Weapon Systems: What's the problem? Uppsala University – Department of Theology.

Bench-Capon, T., & Modgil, S. (2017). Norms and value based reasoning: Justifying compliance and violation. *Artificial Intelligence & Law Review, 25*, 29–64

Benjamins, V. R. & García I. S. (2020). Towards a framework for understanding societal and ethical implications of Artificial Intelligence. *Vulnerabilidad y cultura digital* by Dykinson. pp 87–98.

Black, J. (2002) Critical reflections on regulation. *Australian Journal of Legal Philosopy, 27*, 1–35. http://www.austlii.edu.au/au/journals/AUJlLegPhil/2002/1.pdf. Accessed 30 July 2020.

Bonnemais, V., Saurel, C., & Tessier, C. (2018). Embedded ethics: Some technical and ethical challenges. *Ethics and Information Technology, 20*, 41–58

Buiten, C. M. (2019). Towards intelligent regulation of artificial intelligence. *European Journal of Risk Regulation, 10*(1), 41–59

Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., Newman, P., Parry, V., Pegman, G., Rodden, T., Sorrell, T., Wallis, M., Whitby, B., & Winfield, A. (2017). Principles of robotics: Regulating robots in the real world. *Connection Science, 29*(2), 124–129

Bonnemains, V., Saurel, C., & Tessier, C. (2018). Embedded ethics: Some technical and ethical challenges. *Ethics Information Technology., 20*, 41

Borgesius, F. Z. (2018). Discrimination, artificial intelligence, and algorithmic decision-making. Study for the Council of Europe.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Bryson, J. J. (2018). Patiency is not a virtue: The design of intelligent systems and systems of ethics. *Ethics and Information Technology, 20*, 15–26

Butterworth, M. (2018). The ICO and artificial intelligence: The role of fairness in the GDPR framework. *Computer Law & Security Review, 34*, 257–268

Calo, M. R. (2011). Peeping hals. *Artificial Intelligence Review, 175*, 940–994

Calo, M. R. (2015). Robotics and the lessons of cyberlaw. *California Law Review, 103*(3), 513–563

Caron, M. S., & Gupta, A. (2020). The social contract for AI. Cornell University. https://arxiv.org/abs/2006.08140v1 Accessed 6 Dec 2020.

Canada Government. (2020). Algorithmic Impact Assessment. https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html. Accessed 15 Dec 2020.

Carter, D. (2020). Regulation and ethics in artificial intelligence and machine learning technologies: Where are we now? Who is responsible? Can the information professional play a role? *Business Information Review., 37*(2), 60–68

Cath, C., Watcher, S., Mittelsadt, B., Taddeo, M., & Floridi, L. (2017). Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. https://ssrn.com/abstract=2906249 or https://doi.org/10.2139/ssrn.2906249. Accessed 21 June 2019.

Cave, S., Nyrup, R., Vold, K., & Weller, A. (2019). Motivations and risks of machine ethics. *Proceedings of the IEEE, 107*(3), 562–574

Cerka, P., Grigiene, J., & Sirbikite, G. (2015). Liability for damages caused by artificial intelligence. *Computer Law & Security Review, 31*(3), 376–389

Cerka, P., Grigiene, J., & Sirbikyte, G. (2017). Is it possible to grant legal personality to artificial intelligence software systems? *Computer Law & Security Review, 33*(5), 685–699

Conitzer, V., Sinnott-Armstrong, W., Borg, J. S, Deng, Y., & Kramer, M. (2017). Moral decision making for artificial intelligence. *AAAI Publication, 31º Conference on Artificial Intelligence*

Council of Europe. (2018). European commission for the efficiency of justice, 'European ethical charter on the use of artificial intelligence in judicial systems and their environment. https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c. Accessed 30 July 2020.

Davis, E. (2015). Ethical guidelines for a superintelligence. *Artificial Intelligence Review, 220*, 121–124

Dennis, L., Fisher, M., Slavkovik, M., & Webster, M. (2016). Formal verification of ethical choices in autonomous systems. *Robotics and Autonomous Systems, 77*, 1–14

Djeffal, C. (2018). Sustainable AI Development (SAID): On the road to more access to justice. https://ssrn.com/abstract=3298980 or https://doi.org/10.2139/ssrn.3298980. Accessed 30 July 2020.

Donahoe, E., & Metzger, M. M. (2019). Artificial intelligence and human rights. *Journal of Democracy, 30*(2), 115–126

Dubai (2019). Smart Dubai. Artificial intelligence principles and ethics. https://smartdubai.ae/initiatives/ai-principles-ethics. Accessed 20 July 2020.

EU GDPR. (2016). European Parliament. General Data Protection Regulation. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679. Accessed 30 July 2020.

EU Parliament. (2012). Charter of Fundamental Rights of the European Union (2012/C 326/02), *Official Journal of the European Union*, 2012 C 326, (pp. 391).

European Commission. (2019). Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the Regions. Brussels. https://www.eea.europa.eu/policy-documents/communication-from-the-commission-to-1. Accessed 30 July 2020.

Firth-Butterfield, K. (2017). Artificial Intelligence and the Law: More questions than answers. *Scitech Lawyer, 14*, 28–31

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. C., & Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI. Berkman Klein Center for Internet & Society.

Floridi, L., Cowls, J., King, T., & Taddeo, M. (2020). How to design AI for social good: seven essential factors. *Science and Engineering Ethics, 26*, 1771

French, P. M. (2018). For a Meaningful Artificial Intelligence: Toward a French and European Strategy. Mission assigned by the French Prime Minister. https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf. Accessed 30 July 2020.

Future of Life Institute. (2019a). National and International AI Strategies. https://futureoflife.org/national-international-ai-strategies/. Accessed 20 September 2019.

Future of Life Institute. (2019b). Ansilomar AI Principles. https://futureoflife.org/ai-principles/. Accessed 20 September 2019.

German Federal Government. (2018). German Federal Ministry of Education and Research, the Federal Ministry for Economic Affairs and Energy, and the Federal Ministry of Labor and Social Affairs. Artificial Intelligence Strategy. https://www.ki-strategie-deutschland.de/home.html. Accessed 30 July 2020.

Greene, D., Hoffmann, A. L., & Stark, L. (2019). Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. *Hawaii International Conference on System Sciences* 52nd, 2019.

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly, 30*(3), 611–642

Gurkaynak, G., Yilmaz, I., & Haksever, G. (2016). Stifling artificial intelligence: Human perils. *Computer Law & Security Review, 32*(5), 749–758

Hagendorff, T. (2019). The ethics of AI ethics: An evaluation of guidelines. CoRR, abs/1903.03425.

Hilb, M. (2020). Toward artificial governance? The role of artificial intelligence in shaping the future or corporate governance. *Journal of Management and Governance.*

Hildebrandt, M. (2018). Algorithmic regulation and the rule of law. *Philosophy Transactions of the Royal Society,* 376 (2128).

Holder, C., Khurana, V., Harrison, F., & Jacobs, L. (2016a). Robotics and law: Key legal and regulatory implications of the robotics age (Part I of II). *Computer Law & Security Review, 32*(3), 383–402

Holder, C., Khurana, V., Hook, J., Bacon, G., & Day, R. (2016b). Robotics and law: key legal and regulatory implications of the robotics age (Part II of II). *Computer Law Secure Review, 32*, 557–576

House of Lords. (2018). AI in the UK: Ready, willing and able? *Select Committee on Artificial Intelligence*, Report of Session 2017–19. 13 March 2018.

IEEE. (2019). Ethically Aligned Design. Committees of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. 2nd version. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf. Accessed 20 July 2020

IEEE. (2020). a call to action for business using AI—Ethically aligned design for business. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead/ead-for-business.pdf. Accessed 20 July 2020.

Jackson, B. W. (2019). Artificial Intelligence and the Fog of Innovation: A deep-dive on governance and the liability of autonomous systems. 35 *Santa Clara High Tech*. L.J. 35

Jackson, B. W. (2020). Cybersecurity, privacy, and artificial intelligence: An examination of legal issues surrounding the European Union General Data Protection Regulation and Autonomous Network Defense, 21 *Minnesota Journal of Law, Science & Technology*, *21*

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2Accessed20July2020

Japanese Cabinet Office. (2019). Social principles of human-centric artificial intelligence. Council for science, technology and innovation, https://www8.cao.go.jp/cstp/english/humancentricai.pdf. Accessed 20 July 2020

Kaal, W. A., & Vermeulen, E. P.M. (2017). How to regulate disruptive innovation: From facts to data. *Jurimetrics*, *57*(2).

Kozuka, S. (2019). A governance framework for the development and use of artificial intelligence: Lessons from the comparison of Japanese and European initiatives. *Uniform Law Review, 24*, 315–329

Kunz, J. (1949). The United Nations declaration of human rights. *American Journal of International Law, 43*(2), 316–323

Larsson, S. (2020). On the governance of artificial intelligence through ethics guidelines. *Asian Journal of Law and Society*, 1–23.

Lenardon, J. P. A. (2017). The Regulation of Artificial Intelligence. *Master Thesis. Tilburg Institute for Law, Technology and Society*. Netherlands.

Lewis, D., Hogan, L., Filip, D., & Wall, P. J. (2020). Global challenges in the standardization of ethics for trustworthy AI. https://doi.org/10.5281/zenodo.3516525. Accessed 30 July 2020.

Lamo, M. & Calo, R. (2018). Regulating Bot Speech. *UCLA Law Review 2019*, July 16, 2018.

Leitner, C., & Stiefmueller, C. M. (2019). Disruptive technologies and the public sector: The changing dynamics of governance. In A. Baimenov & P. Liverakos (Eds.), *Public service excellence in the 21st century.* (pp. 238–239). Palgrave Macmillan.

Lewis, T., & Yildirim, H. (2002). Learning by doing and dynamic regulation. *The RAND Journal of Economics*, *33*(1), 22–36. www.jstor.org/stable/2696373 Accessed 20 July 2020.

Lin, P., Abney, K., & Bekey, G. (2011). Robot ethics: Mapping the issues for a mechanized world. *Artificial Intelligence Review, 175*, 942–949

Lin, Y., Hung, T., & Huang, L. T. (2020). Engineering equity: How AI can help reduce the harm of implicit bias. *Philosophy & Technology*. https://doi.org/10.1007/s13347-020-00406-7

Liu, H. (2017). Irresponsibilities, inequalities and injustice for autonomous vehicles. *Ethics Information Technology Journal, 19*, 193–207

Maluf S. (1995). *Teoria Geral do Estado*. 23ª ed., 205–208. Editora Saraiva. São Paulo.

Mantelero, A. (2018). AI & Big Data: A blueprint for human rights, social and ethical impact assessment. *Computer Law & Security Review, 34*(4), 754–772

Mika, N., Nadezhda, G., Jaana, L., & Raija, K., (2019). Ethical AI for the governance of the Society: Challenges and opportunities. *CEUR Workshop Proceedings*, *2505*, 20–26. http://ceur-ws.org/Vol-2505/paper03.pdf. Accessed 20 July 2020.

Millar, J. (2016). An ethics evaluation tool for automating ethical decision-making in robots and self-driving cars. *Applied Artificial Intelligence, 30*(8), 787–809

Monetary Authority of Singapore. (2019). Monetary Authority of Singapore. Principles to promote fairness, ethics, accountability and transparency (FEAT) in the use of artificial intelligence and data analytics in Singapore's Financial Sector. https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/

Monographs%20and%20Information%20Papers/FEAT%20 Principles%20Final.pdf. Accessed 20 July 2020
Nevejans, N. (2016). European civil law rules in robotics. Study requested by the European Parliament's Committee on Legal Affairs. *Policy Department Citizens' Right and Constitutional Affairs*.
Neznamov, A. V. (2020). Regulatory landscape of artificial intelligence advances in social science, education and humanities research, *volume* 420 pp 201–204. XVII *International Research-to-Practice Conference 2020*. Atlantatis Press.
Organisation for Economic Co-operation and Development (2019). 'Recommendation of the Council on Artificial Intelligence'.
Partnership on AI to Benefit People and Society. (2016) https://www.partnershiponai.org/about/. Accessed 12 July 2020.
Pedro, A. P. (2014). Ética, moral, axiologia e valores: confusões e ambiguidades em torno de um conceito comum. *Kriterion*, vol. 55. Belo Horizonte, nº 130, Dez./2014, 483–498.
Poel, I. V. (2016). An ethical framework for evaluating experimental technology. *Science and Engineering Ethics, 22*(3), 667–686
Prakken, H. (2017). On the problem of making autonomous vehicles conform to traffic law. *Artificial Intelligence & Law, 25*, 341–363
Rahwan, I. (2017). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology., 20*, 5–14
Reed, C. (2018). How should we regulate artificial intelligence? *Philosophy Transactions of the Royal Society, 376*, 2128
Riek, L. D., & Howard, D. (2014). A code of ethics for human-robot interaction profession proceedings of we robot, 2014. SSRN: https://ssrn.com/abstract=2757805. Accessed 20 July 2020.
Rousseau, J. (2016). *The Social Contract*. (202–230). ISBN: 978911495741. London: Sovereign.
Russell, S., & Norvig, P. (1995). *Artificial Intelligence. A Modern Approach*. (pp. 4–5). Prentice Hall.
Scherer, M. U. (2016). Regulating artificial intelligence systems: Risks, challenges, competences and strategies. *Harvard Journal of Law & Technology, 29*(2), 354–398
Schrader, D., & Ghosh, D. (2018). Proactively protecting against the singularity: Ethical decision making AI. *IEEE Computer and Reliability Societies Review, 16*(3), 56–63
Smuha, N. A. (2020). *Beyond a human rights-based approach to AI governance: Promise*. Philosophy & Technology.
Taddeo, M., & Floridi, L. (2018). How AI can be a force for good: An ethical framework will help to harness the potential of AI while keeping humans in control. *Science Review, 361*(6404), 751–752
Toronto. (2020). The Toronto declaration: Protecting the right to equality and non-discrimination in machine learning systems. https://www.torontodeclaration.org/. Accessed 20 July 2020
Tutt, A. (2017). An FDA for algorithms. *Administrative Law Review, 69*(83), 83–123
UK Government. (2018). Government response to House of Lords Artificial Intelligence Select Committee's Report on AI in the UK: Ready, Willing and Able? https://www.gov.uk/government/publications/ai-in-the-uk-ready-willing-and-able-government-response-to-the-select-committee-report. Accessed 31 December 2020
United Nations. (2011). *UN guiding principles on business and human rights*. (p. 2011). UN Human Rights Council.
University of Montreal. (2018). Montreal Declaration for a Responsible Development of Artificial Intelligence. https://www.montrealdeclaration-responsibleai.com/the-declaration Accessed 20 July 2020
US Congress. (2019). H.Res.153 - Supporting the de7velopment of guidelines for ethical development of artificial intelligence. https://www.congress.gov/bill/116th-congress/house-resolution/153?q=%7B%22search%22%3A%5B%22ARTIFICIAL+INTELLIGENCE%22%5D%7D&s=2&r=4
US Congress. (2020). s.3891 – Advancing Artificial Intelligence Research Act of 2020. https://www.congress.gov/bill/116th-congress/senate-bill/3891?q=%7B%22search%22%3A%5B%22ARTIFICIAL+INTELLIGENCE%22%5D%7D&s=3&r=7
Villaronga, E. F., & Heldeweg, M. (2018). Regulation, I presume? Said the robot: Towards an iterative regulatory process for robot governance. *Computer Law & Security Review*, 21 June, 2018.
Verheij, B. (2016). Formalizing value-guided argumentation for ethical systems design. *Artificial Intelligence & Law Review, 24*(4), 387–407
Yeung, K., Howes, A., & Pogrebna, G. (2019). AI governance by human rights-centred design, deliberation and oversight: An end to ethics washing (June 21, 2019). Forthcoming in M Dubber and F Pasquale (eds.) *The Oxford Handbook of AI Ethics*, Oxford University Press (2019), https://doi.org/10.2139/ssrn.3435011. Accessed 15 December 2020.
Wallach, W., & Marchant, G. E. (2018). An agile ethical/legal model for the international and national governance of ai and robotics. *Association for the Advancement of Artificial Intelligence*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6191666/. Accessed 20 July 2020
Waser, M. (2015). Designing, implementing and enforcing a coherent system of laws, ethics and morals for intelligent machines (including humans). *Procedia Computer Science, 71*, 106–111
Wright, S. A., & Schultz, A. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. *Business Horizons, 61*(6), 823–832

Publication 3 : The Attraction of Contributors in FOSP

Contents lists available at SciVerse ScienceDirect

Journal of Strategic Information Systems

journal homepage: www.elsevier.com/locate/jsis

The attraction of contributors in free and open source software projects

Carlos Santos [a,*], George Kuk [b], Fabio Kon [c,1], John Pearson [d,2]

[a] University of Brasilia, Department of Management, Caixa-Postal: 4320, 70910-900 Brasilia, DF, Brazil
[b] Nottingham University Business School, Nottingham, UK
[c] Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, 207-C, Rua do Matão, 1010, Cidade Universitária, 05508-090 São Paulo, SP, Brazil
[d] Department of Management Information Systems, College of Business, Rehn Hall, Southern Illinois University at Carbondale, Rehn 210-A, Mail Code 4627, Carbondale, IL 62901, USA

ARTICLE INFO

Article history:
Received 25 October 2010
Received in revised form 11 July 2012
Accepted 31 July 2012
Available online xxxx

Keywords:
Attractiveness
Open source
Free software
Preferential attachment
Contributors
Contributions
Software development

ABSTRACT

As firms increasingly sanction an open sourcing strategy, the question of which open source project to undertake remains tentative. The lack of established metrics makes it difficult to formulate such strategy. While many projects have been formed and created, only a few managed to remain active. With the majority of these projects failing, firms need a reliable set of criteria to assess what makes a project appealing not only to developers but also to visitors, users and commercial sponsors. In this paper, we develop a theoretical model to explore the contextual and causal factors of project attractiveness in inducing activities such as source code contribution, software maintenance, and usage. We test our model with data derived from more than 4000 projects spanning 4 years. Our main findings include that projects' set of conditions such as license restrictiveness and their available resources provide the context that directly influence the amount of work activities observed in the projects. It was also found that indirect and unintended contributions such as recommending software, despite of being non-technical, cannot be ignored for project activeness, diffusion and sustainability. Finally, our analysis provide evidence that higher attractiveness leads to more code-related activities with the downside of slowing down responsiveness to address projects' tasks, such as the implementation of new features and bug fixes. Our model underscores the significance of the reinforcing effects of attractiveness and work activities in open source projects, giving us the opportunity to discuss strategies to manage common traps such as the liability of newness. We conclude by discussing the applicability of the research model to other user-led initiatives.

1. Introduction

Free and open source software projects (FOSPs) comprise groups of developers and users geographically dispersed but connected together through shared values and the Internet (Herbsleb and Mockus, 2003; Stewart and Gosain, 2006). Open source developers have traditionally developed software as a hobby but are increasingly being paid and sponsored by commercial and public organizations (Fitzgerald, 2006). To facilitate the development process and promote widespread adoption, the application and its source code are made available on a website, which provides all the information and tools

* Corresponding author. Address: Universidade de Brasília, Campus Darcy Ribeiro, Departamento de Administração, Asa Norte – ICC Norte – 1° Andar, Caixa-Postal 4320, 70910-900 Brasilia, DF, Brazil. Tel.: +55 61 82126224.
E-mail addresses: carlosdenner@unb.br (C. Santos), george.kuk@nottingham.ac.uk (G. Kuk), kon@ime.usp.br (F. Kon), jpearson@business.siuc.edu (J. Pearson).
[1] Tel.: +55 11 3091 6135.
[2] Tel.: +1 618 453 7802.

0963-8687/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jsis.2012.07.004

needed for the software to be used, adapted and improved by the public. Over the years, several projects, such as the Web server Apache, the operating system GNU/Linux, and the browser Firefox, have become widely adopted, demonstrating the viability of the open source production model and its capacity to create high-quality applications. Consequently, corporations have started opensourcing their software assets, aiming to create and capture new business value from this alternative to a more traditional way of developing software (Agerfalk and Fitzgerald, 2008).

The significance of attracting users and developers has been frequently highlighted in the open source literature (Arakji and Lang, 2007; Koch, 2004; Krishnamurthy, 2002; Sen et al., 2008; von Krogh et al., 2003). Each group of users and developers contributes a unique set of complementary resources to FOSP: users provide inputs including bug reports, suggestions of new features, and translation of documentation; and developers implement new features, fix bugs, and deal with sponsors. These roles are reflected in the ways the success of FOSP has been measured (Crowston et al., 2005; Long, 2006), including developers' contribution to source code modularity (Shaikh and Cornford, 2003), number of lines of code generated (Mockus et al., 2000), velocity of closing bugs (Herbsleb and Mockus, 2003), and the number of downloads (Crowston et al., 2004; Krishnamurthy, 2002; Grewal et al., 2006).

Raja and Tretter (2006), Crowston and Scozzi (2002), and Comino et al. (2007) viewed success as the ability of a project to advance through development phases (e.g., from alpha to beta, and from beta to stable). Koch (2004) and Crowston and Howison (2006) suggested the use of community size (i.e., number of members) as a proxy for success. Additionally, Stewart and Gosain (2006) adopted a dependent variable labelled "effectiveness", composed by the abilities to receive inputs and produce related outputs such as fixing bugs and adding new features to the software. These various measures reflect the roles of both input (e.g., bug reports) and output (e.g., bug fixes) producers in open source software development and success. In short, contributions to FOSP come from various groups; developers need users to inform their practices, and users need developers to implement their requests (von Krogh and von Hippel, 2006). Yet, the literature primarily singles out developers as the main contributor to the success of FOSP (Bagozzi and Dholakia, 2006).

The eye-ball metaphor, underpinning the Linus law, underscores two sources of contributions for the success of FOSP. Besides reviewing each others' source code (co-developers), it highlights the role of users as beta-testers in reporting bugs (Raymond, 1999). Bagozzi and Dholakia (2006) noted that experienced users provide support to less experienced individuals, and Bevan (2006) pointed out that users' input is frequently responsible for usability improvements. Users play an important role in the innovation process of FOSP. Although not generally acknowledged in empirical research, users' contributions are critical to FOSP success (Crowston et al., 2003; von Hippel, 2005; Grewal et al., 2006; Bagozzi and Dholakia, 2006), and their contributions have been observed in related industries (Arakji and Lang, 2007).

Another source of contribution is the role of the visitor to FOSP's Web pages. We suspect that this type of contribution is less frequent, but visitors contribute to FOSP through various activities such as reporting broken links or installation problems, and requesting a version not yet available for a particular operating system (or a missing feature that they wish to have). There is also the possibility of a technically-inclined visitor inspecting the source code and posting a suggestion or referring people to the project. Hence, visitors can contribute to FOSP success, even though they may never use or get directly involved in developing the software.

Ye and Kishida (2003) discuss the roles of passive users, readers, bug reporters, bug fixers, peripheral developers, active developers, core members, and project leaders. Though fairly comprehensive, their model implicitly assumes that the contributing roles can only be performed by users and developers, excluding the visitors as a key actor. We expand this view by stating that visitors, users and developers are key resources to FOSP, and that attracting, retaining and inducing them to contribute are core challenges for success. Yet how their roles are interrelated in open source development is relatively unexplored.

To address this limitation, we introduce a theoretical framework around the construct of "attractiveness", which is defined as an array of project values perceived by its potential and actual visitors, users and developers. Our motivation to define and focus on this construct is based on prior research, which has suggested "the need for the company to market the attractiveness of the project and improve its visibility" (Agerfalk and Fitzgerald, 2008, p. 394). This definition implies that attractiveness is a core construct in obtaining the critical resources brought along by different actors, whose motivations, capabilities and likelihood to contribute are influenced by their perceptions of project values and the available resources to the project. In turn, their contributions may increase project values (such as attractiveness) and perpetuate software development as a virtuous cycle. We argue that contributors, including visitors, users and developers contribute in different ways to the shared-innovation process, and that as a collective promotes wide adoption and software development and maintenance. In developing our model, we seek to identify factors that affect the distribution of these resources and ultimately the relative success among FOSP.

This paper presents a model to examine what attracts contributors, creating an environment likely to improve and promote the project sustainably. We attribute the success of FOSP to a user-driven process, involving complementary contributions from various user groups. Overall, our intent is threefold: to understand what makes certain open source projects preferable for usage and improvement; to tease out the rich-get-richer effect; and notably to provide managers the insights into formulating a strategy to compete for external collaborators and complementors. We organize the rest of the paper as follows. First, we use the extant literature of FOSP to frame and develop our model of attractiveness. Next, we describe the methods and present the empirical testing of our model, which is followed by the results and the implications for theory and practice. Finally, our view on future research and the broader conclusions close the paper.

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

2. Literature review and model development

What makes certain FOSP preferable to others? Prior research concerns the general appeal of a specific project to developers but relatively less to other user groups. As commercial organizations increasingly sanction an open source strategy (Agerfalk and Fitzgerald, 2008; Fitzgerald, 2006), the user uptake will determine the level of sponsorships (Stewart et al., 2006; West and O'Mahony, 2005), and the sustainability of FOSP. This reflects in the strategic thinking of several large open source coalitions such as the Open Source Initiative and the Linux Foundation. One of their core goals is to encourage the development process to be more liberal, by being less restrictive in the terms of use and commercialization potentials of the derivatives, and attract commercial contributions and exploitation. This highlights the significance of bringing developers and different user groups closer together as a collective rather than a collection of disparate entities.

To release the source code to the public creates opportunities for independent and enterprise developers to co-develop with peers and users. This open approach has proved to be successful and exerted a profound effect on the software industry. Yet, despite the highly cited use cases such as GNU/Linux, only few hundreds FOSP have managed to succeed (Krishnamurthy, 2002; Xu and Madey, 2004; Koch, 2004). With the vast majority failing, it seems that the action of releasing source code alone is insufficient to attract and sustain contribution. This uneven distribution of contributors conforms to the behavioural phenomenon of preferential attachment commonly observed in online communities including open source (Xu and Madey, 2004).

Preferential attachment has been used to explain uneven distribution of resources among objects of systems as diverse and complex as those associated with the Internet, neural networks and academic authorships. It is often used synonymously with the rich-get-richer effect, a characteristic of the Pareto or heavy-tailed distribution (Price, 1976; Clauset et al., 2009; Papadakis and Tsionas, 2010). The tendency of overly concentrating on a few gives rise to a general scale-free, power-law distribution of resources to objects (Simon, 1955), which partially defines the growth mechanism of how such systems develop over time (Barabási and Albert, 1999; Barabási, 2005).

Although the parsimonious and general description offered by preferential attachment is intuitively appealing, it is not clear which set of conditions instigate such distribution. Although prior studies have shown that new nodes tend to connect to highly-connected ones, the underlying reasons (including the contextual conditions) of why certain nodes become established as preferential among other nodes are unexplored. The underlying mechanisms of preferential attachment are markedly different across social and physical systems (e.g., Lee et al., 2006; Newman, 2002). The reasons leading to the concentration of contributors on a few projects are not the same nor directly comparable to physiological systems. In physiological systems, the chemistry guides the interaction among proteins and the attraction is often determined at the molecular level, whereas in social systems the attraction is wired to perceptions and intentions. In relation to open source software development, a patch of code is being reused by developers because of its perceived quality and usefulness. This raises further questions of whether the attachment of new resources to a particular object is solely based on the object's "popularity" and/or its unique set of attributes or characteristics. Additionally, little is known of the underlying change process that can elevate the status of a less preferred object. This paper seeks to examine the set of project conditions and the underlying dynamics of what makes an open source project attractive, aiming to inform practice and generate a theory designed to the open source ecosystem, a kind of theory that, being specific, has been overlooked (Keller, 2005).

The success of open source software has been attributed to many developers working under the Bazaar paradigm. However, only a few FOSP managed to build virtuous and productive Bazaar ecosystems (Krishnamurthy, 2002), giving rise to a scale-free network (Xu and Madey, 2004). Researchers have been trying to understand what motivates FOSP developers and drives user-interest, highlighting the importance of trust, knowledge sharing, employment prospects, sponsorship, coordination mechanisms and communication patterns within the communities (von Krogh, 2002; Crowston and Scozzi, 2002; Stewart et al., 2006; Stewart and Gosain, 2006; Crowston and Howison, 2006; Fershtman and Gandal, 2007; Fang and Neufeld, 2009). Nevertheless, most of the prior research has focused on developers to explain projects' activities and success, and less on the role of users and visitors as contributors to the use-value of FOSP.

Users provide relevant problems to be solved such as bug reports and enable the network externalities that not only increase the popularity of the project but also attract new users and sustain developers' contribution. Similarly, visitors of a project Web page, representing "brand" exposure and commercial success (Grewal et al., 2006), can contribute to the network externalities by enhancing the ranking of the visited project (Muffato, 2006). Moreover, visitors may indirectly contribute to the improvement of FOSP by reporting broken links or engaging in R&D activities. Visitors, users and developers are valuable resources to FOSP as they all are, directly or indirectly, contributing to the projects in terms of R&D, marketing, and technology adoption, improvement and diffusion. Yet their conjoint role in open source software development has been overlooked.

2.1. Project activities: the sources of improvement, software maintenance

We have stated that visitors, users and developers are critical resources to FOSP because they are responsible for contributions, which in turn, are the sources of open source software development, improvement and diffusion (Ye and Kishida, 2003). These contributions can be observed through FOSP activities that take place over supporting online tools, such as forums and bug tracks. We refer to project work activities as the inputs as well as outputs that the community provides to the project.

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

FOSP attractiveness influences work activities in a variety of ways. First, from the perspective that every problem is obvious to someone in software development (Raymond, 1999; Sharma et al., 2002), a straight forward implication of having more visitors, users and developers is that the probability of receiving contributions (inputs and outputs) increases as more people gravitate around the project. Second, from a theoretical point of view, we argue that project attractiveness influences the community motivation, at the individual level, to contribute to the project. Attractiveness is related to popularity and visibility of a project, which increases the motivation of individuals to showcase their abilities (signalling), and improving their reputation within both developers and business communities (Lerner and Tirole, 2002; Roberts et al., 2006). In short, the development community may expect a higher impact from their contributions, as well as higher returns, and is more inclined to contribute to projects they perceive as highly attractive.

Free software projects, as creative enterprises, have agents embedded in an open ecosystem that can inspire and evaluate their contributions and resulting products (Guimera et al., 2005). This "large social milieu", as Nonneke and Preece (2000, p. 6) put, is a source of motivation to contributors and "has far-reaching consequences", affecting people's posting behaviour. The richer this ecosystem, or larger the social milieu, the better for the project, as it becomes more diverse, fostering innovation (von Hippel and von Krogh, 2003; O'Mahony, 2007). A highly-attractive project that brings a sufficient number of developers and users/visitors together has a higher chance of forming a virtuous, cooperative relationship with the users and visitors sourcing a relevant set of problems for the developers to solve. Higher user-interest leads to more development activity (Stewart et al., 2006). In contrast, the use-value of a project without visitors and users is limited to a few developers and has a lesser appeal to the wider public, particularly affecting community intention to contribute.

To explore these theorised benefits of attractiveness on FOSP activities, we focused on four complementary, yet different, measures. First, intending to capture the amount of any direct activity observed in the projects, we gathered and summed the numbers of bug reports, feature and support requests, and patches submitted, under the label of project "activeness". This measure focuses on the volume of ideas and opportunities for project improvement and maintenance that were suggested by its resources. Then from this total sum, we excluded any requests that the project was unable to address, maintaining only those that were properly taken care of ("closed"). We labelled this second measure "effectiveness", and believe it is key for the long-term success of a project, as its absence would condemn an open source software to an outdated state, progressively distant from market's changing interests and demands.

Further, we calculated the ratio of effectiveness over activeness to explore the effects of attractiveness on the likelihood of a contributor input being properly addressed. This represents a project's overall responsiveness to tasks originated in the community. We also computed the average time projects take to address the inputs they have received. The importance of development speed is practical and generates some level of urgency among developers, in that, "[t]he more readily developers can recognize the needs and problems addressed by the project, the more successful the project" (Crowston and Scozzi, 2002, p.10). Details of the construction and acquisition of these measures are discussed in the methods and results sections.

Having discussed the main project activities related to software maintenance and improvement, we argue that there is a cyclical influence between project activities and attractiveness. Resources, influenced by attractiveness, are recruited and act to maintain and improve software, and their recruiting and actions influence project attractiveness. At an empirical level, we will first assess whether there is a direct influence from attractiveness to project activities, a necessary condition to support our theoretical claim of cyclicality. To do that, we formulate our first four propositions, linking attractiveness to project activities. Accordingly, we have:

Proposition 1: FOSP attractiveness significantly influences activeness.
Proposition 2: FOSP attractiveness significantly influences effectiveness.
Proposition 3: FOSP attractiveness significantly influences likelihood of task completion.
Proposition 4: FOSP attractiveness significantly influences time for task completion.

2.2. The causes of attractiveness

As we have discussed, visitors, users and developers tend "to attach" to few "attractive" open source projects, creating a rich gets richer effect (Krishnamurthy, 2002; Xu and Madey, 2004; Koch, 2004). In the following sections, we focus on specific characteristics of FOSP, which define their "condition" that impacts attractiveness, and relate them later onto the preferential attachment mechanism.

2.3. Set of conditions: FOSP characteristics

To explain involvement in open source development, prior research has focused on contributors' intrinsic and extrinsic motivations, like signalling to potential employees (Crowston and Scozzi, 2002; Stewart and Gosain, 2006). We do not challenge this explanation, but intend to expand it, using project as our unit of analysis (Colazo and Fang, 2009). The focus on project allows the examination of which project set of conditions (a set of contextual factors), influences the likelihood of that project being chosen by potential contributors. Project characteristics have been shown in the past to be linked to contributors' motivations and perceived usefulness, impacting development activities and adoption rates (Comino et al., 2007; Crowston and Scozzi, 2002; Fang and Neufeld, 2009; Sen et al., 2008).

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

The project set of conditions, including license type and application domain, affect their attractiveness throughout their life-cycles. This in turn influences contributor recruitment and contribution generation, which take place as people browse for and find out about software, as well as consider contributing to a project after they have become "a resource of". Some people may report a failed-attempt to download or install the software; others may post bugs or develop features for applications in initial stages. Moreover, open source advocates might prefer to use, be associated with, and/or refer people to GPL-licensed applications. FOSP's application domain helps define the target population size, benefiting those in larger domains. For example, it is highly likely that there is a smaller demand for compilers in comparison to office suites. Accordingly, we expect that, all other things being equal, compilers are less likely to attract contributors than office suites, impacting the recruitment of contributors, project's activity level and its probability of receiving other indirect contributions. Similarly, application domain relates to contributors' profile, as users of compilers tend to be technically-inclined software developers, and thus are more capable of contributing or inspecting source code than office suite users.

A project's set of conditions affects its ability to attract resources and activity levels via different mechanisms. First, these conditions offer an opportunity for people to prefer a specific application context when choosing what to adopt and contribute to. Second, they specify the boundaries of competition for recruiting a limited amount of resources available in the marketplace. Finally, FOSP set of conditions influences their visitors, users and developers likelihood of contributing, as their community profile (e.g., computer skills) depends on who they target (programmers or end-users). These project conditions work together to influence people's perception of FOSP attractiveness, which affects recruitment of resources and generation of contributions.

Several distinctive and empirically observable elements constitute the project set of conditions, working in tandem with each other as a set to influence people's perceptions of project attractiveness in a cyclical and dynamic manner. As such, the task to hypothesize in advance the direction of influence of all their combinations on our variables of interest would be unmanageably complex. Accordingly, we opted to theorize in an exploratory manner, stating that each project condition has an influence, which is not independent of the state of another condition as they act together as "the context", on project's resource availability and the amount of activity observed. Next, we present our propositions relating each FOSP characteristics to attractiveness and work activities, and then elaborate on how feedback works within the preferential attachment mechanism along with project conditions.

2.4. Type of license

What allows the classification of a project as open source and/or free software is the license. FOSP licenses regulate what can and cannot be done with the software, its source code and derivative works, influencing its range of use and distribution, and notably its intellectual property (Agerfalk and Fitzgerald, 2008; Santos et al., 2011). Under the General Public License (GPL) the source code of the new derivatives have to remain open and are not allowed to be redistributed as proprietary software. Whereas the Mozilla Public License and the Eclipse Public License permit greater interaction with proprietary software and provide greater commercial freedom (Fershtman and Gandal, 2007).

In the literature, open source software licenses are commonly grouped based on their restrictiveness (Fershtman and Gandal, 2007; Lerner and Tirole, 2005; Sen et al., 2008; Stewart et al., 2006). In general, there are three levels of restrictiveness: (1) do not allow combined compilation with proprietary software and force a derivative work to have the same license as the original (Strong-Copyleft); (2) force derivative works to have the same license but allow combined compilation with proprietary software (Weak-Copyleft); and (3) do not impose any of these restrictions (Non-Copyleft). The influence of type of license on FOSP has appeared in different ways. Lerner and Tirole (2005) have examined how license choice is associated with a project's audience, developers or end-users. It has been reported that license choice is associated with the amount of developer activity, user interest, and individual intention to contribute (Fershtman and Gandal, 2007; Stewart et al., 2006; Santos et al., 2011). Sen et al. (2008) pointed out that license restrictions impact perceived usefulness and the visibility of software. Colazo and Fang (2009), using social movement theory, argued that license type, FOSP size, and development speed are linked. These findings together suggest that license type affects project work activities and thus attractiveness of open source software. That is, the attraction of resources is influenced by the license a project adopts. For instance, whilst open source advocates tend to use GPL applications, for-profit organizations tend not to combine their proprietary codes with open source ones because this will effectively give away their property rights. Thus, we have:

Proposition 5.1: Type of license significantly influences FOSP attractiveness.
Proposition 5.2: Type of license significantly influences FOSP activeness.
Proposition 5.3: Type of license significantly influences FOSP effectiveness.
Proposition 5.4: Type of license significantly influences FOSP likelihood of task completion.
Proposition 5.5: Type of license significantly influences FOSP time to complete tasks.

2.5. Type of user

Software aid processes that are managed by different types of users. These types of users can be end-users (e.g., browsing the web), advanced end-users (e.g., developing a database), system administrators (e.g., creating backups), and developers (e.g., writing software). The influence of type of user, or audience, on FOSP activities has been discussed in the literature

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

(Crowston and Scozzi, 2002; Stewart et al., 2006; Fershtman and Gandal, 2007). FOSP aiming at technically inclined users are more likely to find contributors among their users, and FOSP for less technically inclined users have a larger audience and, thus, more chances of finding users and developers (Comino et al., 2007; Crowston and Scozzi, 2002). Some FOSP "have a greater number of potential developers in the community than others do", says Johnson (2002, p. 664). Additionally, Stewart et al. (2006, p. 136) hypothesized the influence of type of user on both user-interest and development activities, stating that "[t]hose targeted at a developer audience may attract greater development activity or be less appealing to users". Similarly, projects targeted at developers have been found to be more active than those targeted at system administrators, which outperform projects for end-users (Crowston and Scozzi, 2002). Finally, the number of software developed for different types of users are not equal, influencing the amount of available resources in the market and competitiveness for contributors and their contributions.

Although the literature suggests that recruitment and project work activities are affected by type of user, it is not clear how these will affect the conjoint attraction and role of visitors, users and developers in FOSP, nor in the presence of and interacting with the other FOSP characteristics. We formulate the following:

Proposition 6.1: Type of user significantly influences FOSP attractiveness.
Proposition 6.2: Type of user significantly influences FOSP activeness.
Proposition 6.3: Type of user significantly influences FOSP effectiveness.
Proposition 6.4: Type of user significantly influences FOSP likelihood of task completion.
Proposition 6.5: Type of user significantly influences FOSP time to complete tasks.

2.6. Application domain

An application supports certain processes for its users. To cover these processes, FOSP applications are classified according to domains, or project categories, such as genealogy, payroll, chat, browser and games (Crowston and Scozzi, 2002). The application domain restricts, where a software competes for contributors and their contributions; it is the software industry or niche (Jaisingh et al., 2008). For example, Firefox generally does not compete for users with R, as they operate in distinct arenas. Potential contributors typically look for projects to work on in a specific domain (Johnson, 2002), and so it is likely that they will first select one over the others. Raymond (1999) uses email client projects to illustrate how self-selection works in practice, underlining how FOSP are competing among themselves for contributors and contributions within each application domain.

Additionally, the application domain has been discussed in the context of "technical sophistication" (Comino et al., 2007). As such, more technically sophisticated domains like "compilers" are more likely to receive substantial contributions from its users, as they are developers and thus have the required technical skills (Crowston et al., 2005; Crowston and Scozzi, 2002). In this line of reasoning, there is an overlapping with the effects of the type of user condition we discussed previously. However, it is still important to control for the application domain in addition to the type of user as projects of the same application domain can still target different types of users. For example, one software can provide an intuitive graphical user interface, whereas a different application may require the user to write algorithms and be familiar with a particular language to perform the same tasks through a command-line interface. The software application domain has been hypothesized to be associated with both user-interest and development activity (Stewart et al., 2006). Therefore, FOSP attractiveness and work activities should be affected by their application domain as well as by the characteristics of the other projects operating *within* that domain. Thus, in the context of our model, we have:

Proposition 7.1: Application domain significantly influences FOSP attractiveness.
Proposition 7.2: Application domain significantly influences FOSP activeness.
Proposition 7.3: Application domain significantly influences FOSP effectiveness.
Proposition 7.4: Application domain significantly influences FOSP likelihood of task completion.
Proposition 7.5: Application domain significantly influences FOSP time to complete tasks.

2.7. Stage of development

Software engineers generally classify applications according to their stage of development, in a life-cycle fashion. These stages in the life-cycle are planning, pre-alpha, alpha, beta, production and mature. FOSP make their software stage available to the public, informing their maturity level condition and influencing decisions to adopt and contribute (Crowston and Scozzi, 2002; Raja and Tretter, 2006), as well as the level of development activity (Stewart et al., 2006).

There are known links between stage of development, project size and contributors' technical skills (Comino et al., 2007). Furthermore, stage of development can be used as a strategy by project managers, for a beta release may be seen as an invitation for contributions from the community, as well as for users to try a "new" application. At the same time, it is probable that users prefer mature software over applications at the alpha level. Accordingly, these observations together suggest that the stage of development condition has to be incorporated in our model, influencing how attractive FOSP are and their level of work activities. Thus, we have:

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

Proposition 8.1: Stage of development significantly influences FOSP attractiveness.
Proposition 8.2: Stage of development significantly influences FOSP activeness.
Proposition 8.3: Stage of development significantly influences FOSP effectiveness.
Proposition 8.4: Stage of development significantly influences FOSP likelihood of task completion.
Proposition 8.5: Stage of development significantly influences FOSP time to complete tasks.

2.8. The feedback effect of project activities, software maintenance

FOSP change their contributor-base and the amount of contributions received over time, distinguishing between initial and sustained contribution. Fang and Neufeld (2009) demonstrated how important is for contributors to learn and construct a shared identity with the project to sustain their motivation to participate over time. Roberts et al. (2006) showed how past performance rankings of developers influence their future motivation to contribute. Ye and Kishida (2003) pointed out that there is a co-evolution in FOSP, with contributors' motivations being affected by their contributions to the project, which can result in status promotion. Yet, this stream of research has been mostly restricted to the contributor (developer) as the unit of analysis, not the project. Adding to this literature, we propose a mechanism through which contributions are sustained at the project level. Specifically, we differentiate direct (project activities) from indirect contributions (attractiveness self-reinforcing effect).

One source of FOSP improvement and software maintenance is through the work activities carried out by developers. These activities, which are direct contributions, can translate into changing project attractiveness. For example, visitors, users and developers might have requested a series of new features that were later implemented, directly affecting the perceived usefulness of the software and the potential benefits to new participants in the long-run (Stewart et al., 2006). In contrast, a "buggy" software, with many unreported and thus not addressed problems, will adversely affect project attractiveness. When a community crowds around a project, direct participation and improvements will be visible to the public. This visibility will enhance the value of the project/software, attracting more visitors, users and developers. Notwithstanding, the effects of a project conditions on attractiveness and work activities do not cease to operate, influencing the likelihood of receiving a contribution in any case.

Our perspective on the relationships of project conditions, attractiveness and work activities over time can be summarized illustratively as follows. To begin with, a recently created project has low attractiveness as it is only in the planning stage of development. The only contributors are the project creators and a few of their colleagues, who could test the application after a release. Knowing that, the creators work by themselves, designing the source code structure and defining the application features, writing code and implementing functionalities, developing the user-interface and managing all other project tasks. By doing that and performing what we are calling here project work activities, they release an alpha version of their application, which then improves the attractiveness of the project via changing the initial set of project conditions. Now, the creators can send an email to their colleagues saying that there is a new software for them to test and give feedback. The colleagues download the application and start using it, spotting many problems and realizing that there is a lot more that the application should do in order to be professionally adopted by them. As the colleagues report their impressions, increasing the project index of activeness, the creators have the opportunity to address these relevant issues and improve the project attractiveness even further based on their communication with the users (work activities). If they are able to do so, the application will advance again into the next stage of development, their colleagues will become actual users and feel motivated to report more desired features and/or potential problems. This forms a virtuous cycle, where attractiveness is enhanced both by the maturing application condition, which positions it better against similar applications, and the addressing of users' demands, which affects software quality and project attractiveness, fostering diffusion.

In short, FOSP attractiveness is enhanced through direct contributions to the project, but their set of conditions have an independent influence on attractiveness as well. This proposition is a core assumption of the open source movement and takes the form of a loop-mediation in our model, from attractiveness to project activities and, then, back to attractiveness, controlling for the set of conditions. More formally, we have:

Proposition 9: Past project work activities significantly influence future FOSP attractiveness above and beyond the effects of projects' set of conditions.

2.9. Attractiveness self-reinforcing effect, indirect and unintended contributions

However important, the role of visitors, users and developers in enhancing their projects attractiveness is not restricted to contributing to its software source code or related material, such as website content, support provision and documentation. Contributions to FOSP occur in a variety of ways, including indirect, and even unintended, ones. Stewart et al. (2006), for example, highlighted the contributing role of users via word-of-mouth recommendations, influencing future user-base size and, potentially, developers' intention to contribute, configuring an important indirect contribution. By extension, we expect a similar behaviour from visitors and developers.

To understand the motivations for this behaviour, Nonnecke and Preece (2003, p.126), studying online groups, pointed out that lurkers, people who do not actively contribute (post), by reading daily messages, develop a strong sense of community. This sense of community leads to the dissemination of information, such as "contacting individuals [...] and

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

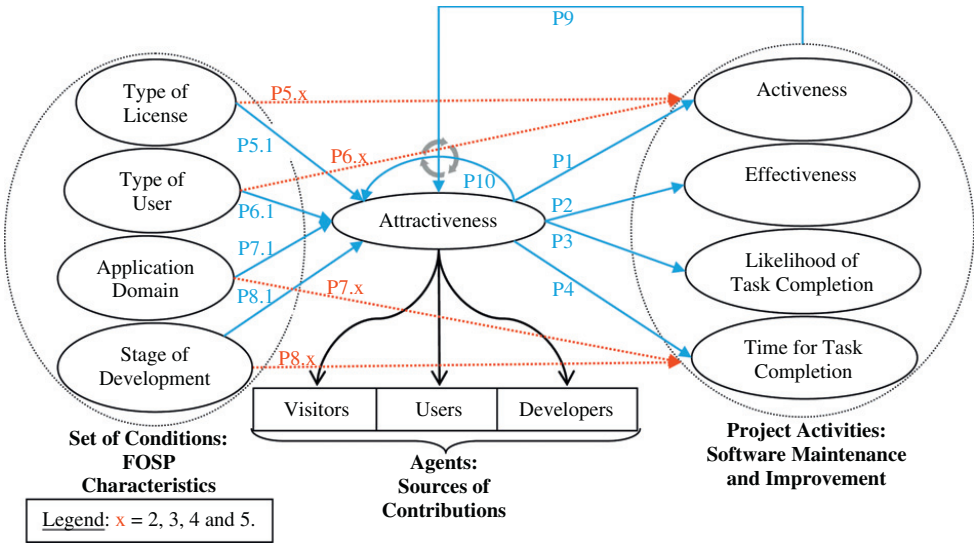


Fig. 1. Theoretical model for the attraction of contributors in FOSP.

introducing others to a group". We find it likely that a similar mechanism takes place in open source projects, as the effort of their contributors have been found to be influenced by common values, beliefs and norms (Stewart and Gosain, 2006). In support to that, Mozilla Firefox's users once contributed their own money to advertise the browser in *The New York Times* (Bagozzi and Dholakia, 2006). Similarly, users may freely create public online content related to open source projects, recommend them to their employers or employees, and wear supporting apparel; visitors can share a project website with their friends on social networks, recommend it to someone by email, donate money, write reviews and rate the project on its repository; and, finally, developers can organize workshops, mentoring programs and "install fests". These represent important, however indirect, contributions to project attractiveness, as they help to build a reputation in the marketplace.

Additionally, there is a more discrete and perhaps unintended type of contribution the community performs. We refer to this as the project density effect, or the silent contribution in Nonneke and Preece's (2000) terms, and its causal chain goes as follows. A visitor, by visiting, and a user, by downloading, enhance a project position in online ranks, increasing its visibility and, thereby, influencing its attractiveness. Especially visitors, by representing website traffic, can help a project find business sponsors, which can bring new developers, and raise money through the sale of ads, which can then be used to publicize the project or hire new developers and sustain work.

All these things together suggest that attractiveness, by affecting the number of available visitors, users and developers and their likelihood of performing indirect contributions, will influence the amount of these sources of contributions in the future, via an effect on attractiveness itself. Again, it is noteworthy that these relationships take place in the presence of projects' set of conditions, which continue to affect attractiveness regardless of an increasing visibility. For example, an organization intending to adopt an open source software may attend a workshop and find out that its license does not allow it to be merged with the organization proprietary applications already in place. Thus, in the context of our model, we formally have:

Proposition 10: FOSP past attractiveness significantly influences future attractiveness above and beyond its effect via project work activities, controlling for projects' set of conditions.

Together, Propositions 9 and 10 represent the specific mechanism through which we claim preferential attachment takes place in open source projects, making the attractive projects even more attractive and leaving the new and unattended project in an difficult situation, struggling to evolve and survive (i.e., liability of newness). With these propositions, we can present of our theoretical model, depicted in Fig. 1, and are now able to move to the discussion of the methods required to evaluate its plausibility, empirically.

3. Methods

The characteristics of our propositions, such as the variety of relationships proposed and the presence of a latent construct, led us to evaluate variables' effects using structural equation modelling (SEM). SEM is a technique that allows researchers to test complex models with simultaneous equations (Chin, 1998; Kelloway, 1998). The structural model developed by us represents a cross-sectional evaluation of variables' theorised influence on the ones that do not represent themselves at a different period in time, effectively testing Propositions 1–8.5 with the data. This model test variables' observed associations at a point in time, a necessary condition for our propositions to find support in the empirical analysis.

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

As Propositions 9 and 10 predict an association of attractiveness from the past with its future value, we decided to develop an independent test of their plausibility. We discuss the methods for testing these propositions later.

The structural model was tested with data from the largest FOSP repository, SourceForge.net. Data about SourceForge.net projects was readily available at the University of Notre Dame, which allowed us to collect three samples (4693 projects in January of 2006, 4500 in 2007, and 4507 in 2008) to compare and improve our confidence in the results. Data was separated into files by period, and projects were filtered out according to the criteria of: (1) not having inadmissible values on the variables, e.g. negative date of creation; (2) having more than one download and member (to increase confidence that the project contains software and source code posted); and (3) not having "inactive" flagged, which indicates that the project is no longer active, or perhaps has never been.

To capture the amount of available resources to a project at a point in time, we used three specific variables: number of hits the project website had as a proxy for number of visitors; the number of times its application was downloaded as a proxy for the number of users; and the number of registered members the project has as a proxy for developers. Next, a factor or component can be extracted from these three measures, which have been found to be significantly and positively correlated with each other in previous studies (Krishnamurthy, 2002). This factor offers a single and straightforward measure of resources' availability to the project.

Nevertheless, these proxies as a composite measure of available resources of a project are not without drawbacks. The use of number of visits as a proxy for visitors, downloads for users, and members for developers is not perfect. The proxies for visitors and developers tend to be biased upward, as a visitor may "hit" a website more than once and a member may never directly develop source code, being in charge of an administrative role such as the assignment of tasks to developers, for example. In its turn, number of downloads tend to bias its representation in a more complex way, as software may be downloaded but never used (upward-bias), or a user may acquire the software through some other form, such as from a GNU/Linux distribution, or even download it more than once (downward-bias). However, one cannot be a visitor without visiting, one of the main ways of becoming a user of open source is downloading the software, and most members are source code developers. And given that we are interested in understanding the distributions of these measures rather than their absolute values, we consider them useful for research purposes. In defence of their usefulness, researchers have used downloads to count users and user-interest (Stewart et al., 2006; Wiggins et al., 2009), and members to represent developers (Fershtman and Gandal, 2011). The FOSP literature frequently adopts these measures (e.g., Crowston et al., 2005; Raja and Tretter, 2006; Stewart and Gosain, 2006; Sauer, 2007).

To explore the role of project work activities and software maintenance, we observed activeness as the sum of (1) bug reports, (2) support requests, (3) feature requests, and (4) patches submitted; effectiveness as the sum of (1)–(4) that were closed (resolved/approved); likelihood of task completion as effectiveness divided by activeness; and time for task completion as the average time a project took to close its tasks.

In accordance with the discussion on the skewed-distribution of resources to objects and previous research on FOSP, hits, downloads and members were found to have skewness values outside the range of normality, and standard deviations much greater than their averages (see Table 1, for the 2008 sample characteristics). Activeness, effectiveness and time for task completion were found to have a similar pattern. Likelihood of task completion was not, but we decided to transform it along with the others that were log-transformed for linearisation to keep the results interpretation consistent. Likelihood was transformed into its inverse-sine-square-root, which tends to render more normally distributed data from variables in the form of proportion (effectiveness divided by activeness) (Crowston and Scozzi, 2002). These transformations for linearisation

Table 1
Descriptive statistics.

	Minimum	Maximum	Mean	Std. deviation	Skewness	
					Statistic	Std. error
Hits	1	1423193	3196.07	32693.98	29.15	.04
Ln.hits	.00	14.17	4.9918	2.32	.16	.04
Downloads	7	160385573	230484.93	2884297.85	41.36	.04
Ln.downloads	1.95	18.89	9.35	2.11	.35	.04
Members	2	374	6.43	9.81	14.89	.04
Ln.members	.69	5.92	1.49	.762	1.01	.04
Activeness	1.00	122375.00	146.38	1874.71	61.67	.04
Ln.activeness	.00	11.71	3.13	1.72	.35	.04
Effectiveness	1	119282	112.02	1811.48	63.29	.04
Ln.effectiveness	.00	11.69	2.49	1.83	.57	.04
Likelihood	.01	1.00	.61	.27	−.24	.04
Arc.likelihood	.01	1.57	.74	.42	.56	.04
Average	21.00	165888141.00	11373856.98	15286540.08	3.40	.04
Ln.average	3.04	18.93	15.30	1.93	−2.2	.04
Life.span	1118.90	3010.00	2111.16	500.50	−.059	.04
Ln.life	7.02	8.01	7.62	.25	−.45	.04
N (listwise)	4507					

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

are frequently reported in open source research (e.g., Comino et al., 2007; Crowston and Scozzi, 2002). We measured the categorical variables, FOSP's set of conditions or characteristics, as dummies.

Sets of dummy variables were created to represent each FOSP condition/characteristic (that is, type of license, type of user, application domain and stage of development). For type of license, we used four dummies, based on Lerner and Tirole (2005): (1) no restriction or non-copyleft; (2) restriction of modification or weak-copyleft; (3) restriction of modification and use or strong-copyleft; and (4) dual-licensing, when a software is licensed on different types of licenses. Type of user required five dummies, application domain required 18, and stage of development 6 (see Table 3). Finally, a control variable was included, life-span measured in days, for its effects on various variables related to FOSP success have been previously reported (Crowston and Scozzi, 2002; Crowston et al., 2005; Stewart et al., 2006; Fershtman and Gandal, 2007).

To generate the results, we used the multisample-SEM capability of EQS 6.1 (Byrne, 2006). Data was entered in its raw form. For the fitting criterion (coefficient estimation), priority should be given to maximum likelihood (ML) when sample size is large (Bentler, 1989; Kline, 1998). Accordingly, we adopted ML, the most common method used by structural modellers (Anderson and Gerbing, 1998; Hair et al., 2006). The results of this statistical analysis are presented after we describe the methods for testing Propositions 9 and 10.

The propositions related to preferential attachment were tested using the process described by Preacher and Hayes (2008) to establish mediation in a single model with multiple mediators and control variables. Their procedure is based on the well-known conceptual description of Judd and Kenny (1981) and Baron and Kenny (1986). However, Preacher and Hayes' procedure has the advantages of testing whether the mediators have an effect as a set and their individual effects in the presence of the other mediators and controls, reducing the likelihood of parameter bias due to the omission of variables (2008). Of course, one should make sure that the mediators included in the model are not conceptually overlapping and highly-correlated. Moreover, Preacher and Hayes' (2008) procedure includes bootstrapping to generate confidence intervals and does not assume normality of the sampling distribution of the indirect effects. According to Preacher and Hayes, these features make the procedure "far superior" to the traditional Sobel test.

To test Propositions 9 and 10, we used data from the Notre Dame 2008 sample and collected additional data of the "future" directly from the SourceForge.net website through a Web crawler our team developed. This new data allowed us to embed some measurement independence in using data of the same variables over time. The final content of the variables in the mediational model is as follows. First, we have the independent variables hits, downloads and members up to 2008. From these three, we calculated one independent variable using the regression weights extracted from the principal component analysis (Mardia et al., 1980) using the statistical package SPSS. Second, the mediators selected for inclusion were effectiveness, likelihood and time for task completion (from the Notre Dame sample). Activeness was left out as it shares a great deal of conceptual overlap with effectiveness, which represents what the community has indeed addressed (source code included, bugs fixed, etc.) and is therefore more interesting to evaluate its effects on attractiveness. Third, there are the dependent variables hits, downloads and members related to the period of January, 2008–December, 2009. Finally, we included all dummies, which represent the projects' set of conditions, and life-span as control variables (covariates). Having gathered all data needed, we used Preacher and Hayes' script (http://www.afhayes.com/public/indirect.sbs) to generate the results, using 5000 bootstrapping resamples to calculate intervals as per their recommendation (Preacher and Hayes, 2008).

4. Results of the structural model

4.1. Descriptive statistics

The Notre Dame SourceForge.net sample is of 149,542 projects in January/2006, 179,867 in 2007, and 143,591 in 2008. After filtering, these numbers were reduced to 4693, 4500 and 4507, respectively. The average project in the 2008 sample received 3196 hits, 230,484 downloads, had six members, and was over 5 years old. Also, the average project produced 146 tasks (inputs received), closing or addressing 112 of them (61%) in an average of 132 days. In their raw form, every variable but life-span and likelihood has a standard deviation greater than the average, and Skewness statistic outside the interval commonly accepted as normal $[-1,1]$. The log-transformations were effective on substantially reducing skewness and returning standard deviations smaller than averages (see Table 1). To illustrate the FOSP characteristics (set of conditions), we noted that the 2008 sample has 1279 projects with licenses that do not impose any restriction to the source code; 2632 are aimed at end-users; 222 were listed under the database application domain; and 1972 projects had their software in the beta stage of development.

4.2. Latent construct reliability

The amount of resources available to FOSP was measured using three variables or indicators and, therefore, its internal consistency had to be assessed. We did so via Cronbach's alpha. The construct scored 0.705 in 2006, 0.711 in 2007, and 0.714 in 2008 (Table 2). Alpha values greater than 0.7 are considered acceptable (Hair et al., 2006; Peterson, 1994; Rutner et al., 2008). Nevertheless, the use of alpha to assess reliability of latent variables is questionable as it requires unrealistically stringent assumptions (Byrne, 2006). Accordingly, we took into consideration EQS reliability coefficient Rho for the overall model as well, which were all above 0.9 and, therefore, indicated appropriate reliability. These results support our claim that

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

Table 2
Model-to-data fit indices for model selection.

	Model with equality constraints			Model without equality constraints			Comparison
	2006	2007	2008	2006	2007	2008	
Sample size	4693	4500	4507	4693	4500	4507	
Cronbach's alpha	0.705	0.711	0.714	0.705	0.711	0.714	
EQS Reliability coefficient rho	0.952	0.955	0.954	0.952	0.955	0.954	
Chi-Square	3042.7 (585 Degrees of Freedom)		2849.4 (237 Degrees of Freedom)			193.3 (348 d.f.)	
P-value for chi-square	<0.01		<0.01			Same	
Model Fit (CFI)	0.979		0.978			0.001	
B–B Normed fit index	0.974		0.976			−0.002	
Root mean square residual	0.021		0.020			0.001	
RMSEA	0.018 – (90% C.I.: 0.017, 0.018)		0.028 – (90% C.I.: 0.027,0.029)			0.010	
			Chi-square critical value (0.05; 348 d.f.):			392.501	
			Decision (given 193.3 < 92.5):			Favour model with constraints	

hits, downloads, and members are likely to have common causes, or are empirical expressions of, at least, one single construct (e.g., attractiveness).

4.3. Overall model-to-data fit

The equation coefficients were calculated both (1) independently, sample by sample and (2) forced to be equal across samples 2006, 2007 and 2008. This strategy was utilized to reduce sampling fluctuations that may obscure effects and bias results (Maitland et al., 2001). The difference in chi-square between the models is 193 (DF = 348; $p > 0.9$). Therefore, the null hypothesis is not rejected, indicating that the more restrictive model (2) produces at least as good a fit as model (1). Also, the two models have similar model-to-data fit indices (Table 2). Thus, the restrictive model (2) is preferable as it is more parsimonious (Mulaik, 2005). Fit tests and indices check if the pattern of covariances are consistent between the specified model and the data (Dow et al., 2008). A "good" fit is a necessary condition to analyse SEM models, that is, CFI greater than 0.95, RMSEA smaller than 0.05, and insignificant chi-square. The constrained model (2) with RMSEA of 0.018, CFI of 0.979 and chi-square of 3042.7 (DF = 585; $p < 0.01$) is acceptable and has good fit except for the chi-square. Nevertheless, the chi-square test is known for its sensitivity to sample size and number of parameters modelled (Kaplan, 2008). Among available fit indices, "RMSEA is relatively [the] most stable" and insensitive to sample size (Yuan, 2005, p. 141). Therefore, we consider the constrained model proper for further analysis.

4.4. Testing the independent effects of variables

The framework developed and coded in EQS for empirical evaluation of Propositions 1–8.5 requires the analysis of five equations. In regression terms, the first equation has attractiveness (F1) as the dependent variable, explained by 33 dummy variables plus life-span (Table 3). As it turned out, life-span is a positive and statistically significant predictor of attractiveness. Thus, the importance of including life-span in FOSP analysis is supported. Also, out of four dummies used to study type of license, one (dual_licensing) was found to influence attractiveness significantly. To register software under licenses with different restrictions has been popular in FOSP with commercial intentions (Santos, 2008; Watson et al., 2008), affecting attractiveness positively. Therefore, we fail to reject P5.1. In general, our rationale to decide whether to reject propositions was that if at least one dummy (e.g., dual_licensing) of a set (e.g., type of license) was significant, then the effect of type of license would have been detected, supporting the proposition.

In relation to type of user, projects for end-users and developers have higher attractiveness, whereas those aiming at others have lower (fail to reject 6.1). Projects listed as multimedia, printing, security and system (application domain) have higher attractiveness, whereas those in database, education, other, scientific and sociology have lower (fail to reject 7.1). Specifically, projects should avoid being listed as others as it hinders attractiveness the most.

Stage of development significantly influences attractiveness in all its six possibilities (fail to reject 8.1). Results indicate that the initial stages of projects (planning, pre-alpha, and alpha) affect attractiveness negatively, whereas advanced stages (beta, production, and mature) enhance attractiveness increasingly, respectively. So, mature projects are more attractive, and to release software in initial stages tend to be ineffective. This first equation explained 22.4% (in 2006), 17.3% (2007), and 15.2% (2008) of attractiveness' variance.

The other four equations indicate how attractiveness influences FOSP work activities. Namely, activeness (F2), effectiveness (F3), likelihood of task completion (F4), and time for task completion (F5). Attractiveness is a significant, and the most important, predictor of the four variables and thus we fail to reject Propositions 1–4. It positively influences activeness and effectiveness, just as Raymond (1999) predicted and many others followed (e.g., Stewart and Gosain, 2006). However, higher attractiveness is associated with smaller likelihood to complete tasks and greater time to complete them. The impact of type of license (both_restrictions) is significant and negative on activeness and effectiveness. Moreover, licenses with both restrictions influence likelihood of task completion positively, suggesting that projects under GPL are more likely to address the

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

Table 3
Structural equations results.

Exogenous variables	F1-Attractiveness 2006, 2007, and 2008		F2-Activeness 2006, 2007, and 2008		F3-Effectiveness 2006, 2007, and 2008		F4-Likelihood of task completion 2006, 2007, and 2008		F5-Time for task completion 2006, 2007, and 2008	
	Coef.	*T*-Statistic	Coef.	*T*-Statistic	Coef.	*T*-Statistic	Coef.	*T*-Statistic	Coef.	*T*-Statistic
	Endogenous variables									
F1-Attractiveness[a]	–	–	3.699	43.586*	3.568	41.946*	−0.208	−14.36*	1.586	25.056*
F2-Activeness[a]	–	–	–	–	–	–	–	–	–	–
F3-Effectiveness[a]	–	–	–	–	–	–	–	–	–	–
F4-Likelihood of Task Completion[a]	–	–	–	–	–	–	–	–	–	–
F5-Time for Task Completion[a]	–	–	–	–	–	–	–	–	–	–
F6-Life-Span[a]	0.172	19.078*	0.006	.161	0.049	1.196	0.031	2.645*	0.72	14.546*
F7-Type of License(No-Restriction)	−0.013	−1.183	0.028	.609	0.065	1.297	0.023	1.557	0.048	0.784
F8-Type of License(Mod-Restriction)	0.01	1.2	−0.041	−1.165	−0.023	−0.593	0.02	1.777	0.02	0.415
F9-Type of License(Both-Restrictions)	0.014	1.622	−0.136	−3.562*	−0.085	−2.036*	0.028	2.254*	−0.026	−0.505
F10-Type of License(Dual-Licensing)	0.027	2.363*	−0.013	−0.254	−0.052	−0.952	−0.028	−1.788	−0.065	−0.983
F11-Type of User(End-Users)	0.093	14.8*	−0.057	−2.179*	−0.092	−3.239*	−0.011	−1.287	−0.036	−1.027
F12-Type of User(Developers)	0.025	3.982*	0.008	0.294	0.025	0.847	0.009	1.082	0.143	3.918*
F13-Type of User(System-Admins)	−0.01	−1.499	0.024	0.858	0.006	0.191	−0.015	−1.708	0.017	0.436
F14−Type of User(Others)	−0.019	−2.227*	0.048	1.336	0.049	1.24	0.011	0.929	−0.145	−3.010*
F15-Type of User(Advanced-End-Users)	0.007	.667	−0.039	−0.944	0.004	0.088	0.054	4.053*	−0.044	−0.783
F16-Application Domain(Communications)	0.001	.109	0.046	1.081	0.004	0.095	−0.039	−2.862*	−0.022	−0.38
F17-Application Domain(Database)	−0.056	−4.898*	0.083	1.693	0.071	1.314	−0.013	−0.842	0.036	0.546
F18-Application Domain(Desktop)	0.019	1.294	−0.129	−2.067*	−0.113	−1.645	−0.013	−0.652	−0.055	−0.652
F19-Application Domain(Education)	−0.059	−4.015*	0.212	3.388*	0.28	4.067*	0.037	1.83	0.158	1.873
F20−Application Domain(Games)	0.003	0.273	−0.157	−3.49*	−0.146	−2.971*	0.017	1.194	−0.027	−0.453
F21-Application Domain(Internet)	0.011	1.323	−0.042	−1.126	−0.018	−0.431	0.034	2.837*	−0.011	−0.216
F22-Application Domain(Multimedia)	0.061	5.124*	−0.234	−4.437*	−0.283	−4.914*	−0.03	−1.817	0.132	1.879
F23-Application Domain(Office)	0.016	1.224	0.256	4.519*	0.232	3.726*	−0.035	−1.914	0.08	1.055
F24-Application Domain(Other)	0.083	−5.437*	0.022	0.334	0.036	0.505	0.006	0.294	0.238	2.731*
F25-Application Domain(Printing)	0.092	3.262*	−0.227	−1.887	−0.181	−1.364	0.041	1.066	−0.138	−0.85
F26-Application Domain(Religion)	0.025	0.644	0.293	1.754	0.328	1.783	0.048	0.894	0.739	3.284*
F27-Application Domain(Scientific)	−0.039	−3.418*	0.043	0.855	0.099	1.783	0.003	0.201	0.109	1.619
F28-Application Domain(Security)	0.066	4.209*	−0.148	−1.088	−0.038	−0.518	0.002	0.084	−0.005	−.054
F29-Application Domain(Sociology)	−0.131	−3.234*	0.585	3.338*	0.537	2.792*	−0.124	−2.222*	0.536	2.278*
F30-Application Domain(Software-Dev)	0.008	1.034	0.02	0.576	0.027	0.718	−0.007	−0.651	−0.036	−0.782
F31-Application Domain(System)	0.035	3.203*	−0.268	−5.544*	−0.33	−6.268*	−0.034	−2.224*	−0.012	−0.184
F32-Application Domain(Terminals)	−0.003	−0.077	−0.148	−.867	−0.213	−1.138	−0.045	−0.822	0.456	1.997*
F33-Application Domain(Text-Editors)	0.025	1.456	0.053	0.684	−0.078	−0.926	−0.062	−2.549*	0.024	0.236
F34-Stage of Development(Planning)	−0.036	−4.181*	0.02	0.556	−0.005	−0.135	−0.027	−2.311*	−0.024	−0.495
F35-Stage of Development(Pre-Alpha)	−0.084	−8.978*	0.034	0.864	0.077	1.772	0.033	2.592*	−0.265	−4.994*
F36-Stage of Development(Alpha)	−0.024	−3.106*	−0.023	−0.694	−0.015	−0.4	0.009	0.802	−0.152	−3.397*
F37-Stage of Development(Beta)	0.03	4.512*	0.12	4.210*	0.16	5.106*	0.007	0.737	−0.048	−1.242
F38-Stage of Development(Production)	0.162	21.72*	0.12	4.026*	0.231	7.053*	0.05	5.248*	0.161	4.031*
F39-Stage of Development(Mature)	0.186	13.961*	0.097	1.746	0.195	3.194*	0.071	3.979*	0.152	2.037*
	Variance explained per sample									
R-Squared	2006; 2007; 2008		2006; 2007; 2008		2006; 2007; 2008		2006; 2007; 2008		2006; 2007; 2008	
	0.224; 0.173; 0.152		0.448; 0.476; 0.49		0.394; 0.414; 0.416		0.028; 0.025; 0.03		0.136; 0.12; 0.112	

[a] Variable log-transformed.
* Significant at 0.05 level; *T*-value >1.96.

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

tasks they received. Finally, no impact of type of license on time for task completion was detected. Thus, we fail to reject P5.2–P5.4, but reject 5.5. Type of user (end-users), which "require extensive and costly usability testing" (Johnson, 2002, p. 656), impacts activeness and effectiveness negatively. Likelihood of task completion is positively influenced by type of user (advanced end-users). And time for task completion is negatively affected by type of user (others) and positively by developers. Thus, we fail to reject Propositions 6.2–6.5.

Application domain affects activeness and effectiveness similarly (positively by education, office, and sociology; and negatively by games, multimedia, and system). However, application domain (desktop) affects activeness negatively, but does not affect effectiveness at all. Application domain (communications, sociology, system, and text-editor) affects likelihood of task completion negatively; whereas Internet do so positively. Finally, the domains religion, sociology, other and terminals tend to take longer to complete tasks. Consequently, we fail to reject P7.2–P7.5. Stage of development (beta and production) influences activeness positively, and beta, production and mature influence effectiveness positively. Moreover, the planning stage influences likelihood of task completion negatively, whereas pre-alpha, production, and mature do so positively. Finally, software in pre-alpha and alpha tend to close tasks faster, and those in production and mature tend to do so slower. All that being consistent with the model proposed, we fail to reject P8.2–P8.5.

Altogether, the results of F2's and F3's equations (activeness and effectiveness) have a very similar pattern when it comes to the significant variables. Among the most interesting results, we found that projects licensed under GPL (strong-copyleft), the most common and restrictive license, tend to be less active as well as less effective than projects that do not adopt GPL. This finding is consistent with previous studies that pointed out that GPL restrictions are seen negatively, decreasing people's intention to contribute (Comino et al., 2007; Fershtman and Gandal, 2007; Lerner and Tirole, 2005), and provides a counter-argument to those who suggested that the fear of open source software being "hijacked" into proprietary applications, maximized by non-restrictive licenses, would drive the community away from contributing (Sauer, 2007; Colazo and Fang, 2009).

A few other comments on the structural model results are worth making. First, software targeted at end-users affects activeness and effectiveness negatively (attractiveness positively), but applications of domains such as education, sociology and office, which are supposedly aimed at end-users too, tend to score higher on activeness and effectiveness (lower on attractiveness). This finding can only be sorted out with a specific study of the interactions between these categories, which can freely vary, the team compositions of these projects, and their user-interfaces. But one way to interpret it is with the logic that these projects have a significant learning curve and thus are not targeted at end-users as we have assumed (e.g., LaTeX). As a matter of fact, 65% of the sociology projects in 2008 were aimed at developers, and only 12% were aimed at end-users solely. The specific purpose of these projects and their characteristics would have to be understood in depth to be sure. Additionally, as the number of projects in sociology, for example, is rather small (17 in 2008), our sample could be biased towards another project condition that is prevalent in the statistical analysis (e.g., out of the 17, 12 are GPL and none is mature).

A second comment on the structural model is that life-span is not a significant influencer of activeness and effectiveness, indicating that the number of inputs and outputs do not increase simply because projects are available for a longer period of time. Likely, the community does not take "seniority" into account to decide whether to contribute to the project, but mainly its attractiveness. Third, the results suggest that more attractive projects have smaller likelihood to complete their tasks, indicating that an overload might occur as more tasks are requested in more attractive projects. In a similar pattern, projects under the GPL license are more likely to complete their tasks, as these projects tend to be less active. However, the variance of likelihood of task completion explained by the model is so low (3%) that, from a practical point of view, its interpretation is limited.

Further, we found that higher attractiveness is associated with more time for task completion, suggesting another side-effect of a higher number of requests, reports and posts (activeness). Having more tasks to deal with and a larger community gathered around these tasks, projects tend to slow down their work-pace, creating a positive chain of influences from attractiveness to activeness to time for task completion. Additionally, higher time to complete tasks may be associated with the type of tasks that are being generated by the community. Projects with more contributors are likely to generate more complex and important tasks to deal with, requiring more time but also being more rewarding. Furthermore, stage of development has an interesting pattern of influence on time for task completion. Projects tend to work faster at pre-alpha and alpha and slower at production and mature. This decrease in activities that accompanies project matureness was predicted by Stewart et al. (2006), and fits well with the logic we discussed before that it is "easier" to contribute in the earlier stages of software development.

The structural model explained 45% of activeness variance in 2006, 47.6% in 2007, and 49% in 2008; 39.4% of effectiveness in 2006, 41.4% in 2007, and 41.6% in 2008; 2.8% of likelihood of task completion in 2006, 2.5% in 2007, and 3% in 2008; and 13.6% of time for task completion in 2006, 12% in 2007, and 11.2% in 2008 (see Table 3). In summary, we failed to reject 23 of 24 propositions, explaining a significant part of FOSP work activities and demonstrating how powerful an understanding of attractiveness can be to manage open software development activities.

5. Results of the mediational model

The final sample used in the mediational model was of 4328 open source projects. The Notre Dame 2008 sample was of 4507, but several projects became inactive from January, 2008 to December, 2009 and, therefore, were excluded. The

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

overlapping sample between the Notre Dame database and the SourceForge.net website, obtained via Web crawler, after applying our filter, was of 4328.

5.1. Principal component analysis: conjoint role of resources

We performed a principal component analysis (PCA) to extract factors from hits, downloads and members from the past (2008) and future (2009). This is a necessary step to perform the mediation analysis based on Preacher and Hayes (2008), which is made using one independent, and one dependent, variable, not a latent construct with three indicators. The parameter we adopted to retain factors from variables was the Kaiser criterion of Eigenvalues greater than one (Stevens, 1986). According to this criterion, both sets of three variables, 2008 and 2009, formed only one factor. The first component extracted from the set of the past had Eigenvalue of 1.97, explaining 65.8% of their variance. Similarly, the set of the future had Eigenvalue of 1.85 with 61.5% of variance explained. No other factor had an Eigenvalue greater than 1. Using the weights of the components extracted, we calculated one variable for each set in a multiple regression fashion.

5.2. The self-reinforcing effects

As it should be consistent with the structural model, in the mediational model (Fig. 2), past attractiveness, controlling for projects' set of conditions, significantly influences effectiveness, likelihood and time for task completion, which, as a set of mediators, affect future attractiveness significantly as well, not allowing the rejection of Proposition 9. The signs of the effects from past attractiveness to each construct related to project activities were also consistent with the structural model results. The total indirect, mediating, effect of past attractiveness through activities on future attractiveness was significant and was calculated by summing the product of their coefficients (i.e., past attractiveness on project activities times project activities on future attractiveness). The bootstrapping resamples provided the intervals to decide for significance. Noteworthy is that although the set of mediators significantly affects future attractiveness indirectly, the direct effect of likelihood of task completion was not found significant in the presence of effectiveness and time for task completion. That is, the interval calculated via bootstrapping for the coefficient of likelihood on future attractiveness includes zero.

To be able to retain Proposition 10 and not reject it, the total effect of past attractiveness on future attractiveness would have to be reduced when controlling for project work activities and set of conditions, but not so much as to bring it to zero. Being reduced to zero would mean that the effect of past attractiveness on future attractiveness is fully mediated by project activities and, therefore, would lead us to the decision of rejecting Proposition 10. The results, in contrast, show that the effect of past attractiveness on future attractiveness is only partially reduced with the inclusion of project activities in the model (from .87 to .8). This means that there is a lot more besides work activities that visitors, users and developers, conjointly, do or represent that affects project attractiveness. The influence of contributors is not restricted to the activities performed via the tools adopted by the project that are publicly available in their repositories.

It is interesting to observe in the mediational model that likelihood of task completion does not affect attractiveness as much as attractiveness influences it. These results suggest that the number of contributors reduces the likelihood of tasks being addressed, but that this operational behaviour does not influence how attractive a project is. Most likely, people are not aware of this before they engage in the project's activities or decide to use it. In addition to that, projects with more contributors take more time to complete tasks and that behaviour does affect attractiveness back in the same direction. A possible explanation for this unexpected finding is that projects that take longer to close their tasks are more careful in doing so or may be working on more substantive issues, which require complex coordination mechanisms and take more time but also reward the project more with future value.

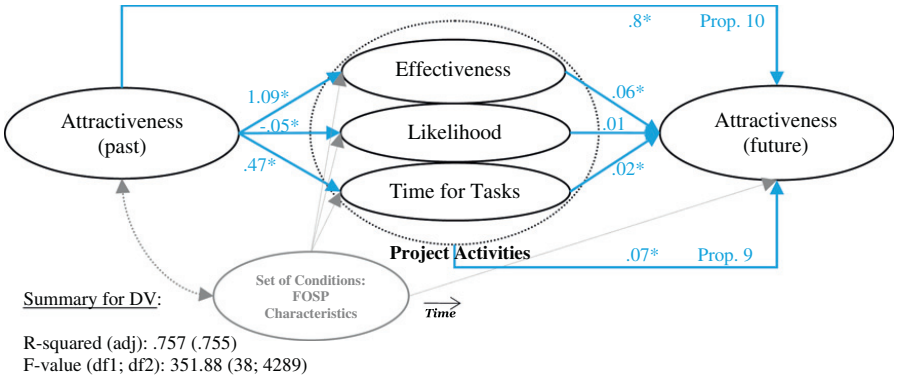


Fig. 2. Results of the mediational model.

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

These results are consistent with our discussion on preferential attachment and the independent and unique self-reinforcing effect of attractiveness, which happens on top of project work activities after controlling for its set of conditions (contextual factors). The mediational model explains about 75% of the variance of future attractiveness.

6. Discussion: Implications for theory and practice

While prior studies have only indirectly stated the importance of attractiveness to FOSP (Agerfalk and Fitzgerald, 2008; Fang and Neufeld, 2009; Sen et al., 2008), this paper seeks to define, further develop and model the attractiveness construct in relation to its causes, contextual factors and impacts, both theoretically and empirically. By arguing that groups of visitors, users and developers provide FOSP with varying and unique sets of development resources that together promote and improve the project towards the sustainability of recruiting new resources and generating more work activities and other indirect contributions, our findings are important to theory and practice in several ways.

Our model elaborates the observed behavioural aspects of preferential attachment specific to open source projects, providing an explanation for how the highly-skewed distribution of resources among them is generated and sustained. Most visitors, users and developers gravitate around only a few projects because of their tendency to select the ''attractive'' project to visit, use, join and contribute, which constitute the key actions on improving a project attractiveness. As this mechanism repeats, a reinforcing cycle is sustained, to the benefit of the already preferred. These selection processes are influenced by FOSP's set of conditions, which not only influences decisions of adopting (directly affecting attractiveness) and contributing (indirectly influencing attractiveness), but also defines projects' competition space for resources, limiting their achievable level of attractiveness. In a hypothetical world, where the population of projects and their set of conditions are stable, resources' stream of actions would be concentrated at predictable levels. However, new projects are created all the time, and old ones change their conditions constantly, disturbing the system of projects to influence resources' decisions and making the task of prediction complex. This dynamic and turbulent environment is what creates opportunities for the new to emerge, and threats for the successful to succumb.

As the results suggest, resources' perception of attractiveness is formed based on: certain contextual factors or characteristics of the project, that is, the identity of an object in comparison to the others; its visibility, which is conditioned by the amount of resources that project has and their indirect contributions; and the work activities performed by their resources towards software maintenance and improvement. Together, project conditions and its members' activities interact to form a dynamic and market-wide reputation.

The tendency to act towards ''the same'' projects set forth their momentum in a reinforcing loop that recursively determines an extreme concentration of resources (making the rich richer). Alternatively, projects can be excluded from resources' decision space for lacking attractiveness and thereby spiral down, making the poor poorer. This process of disturbance could be triggered by a change of sponsors that reconfigures the industry (e.g., OpenOffice in the Sun-Oracle deal), or by a license change (see Santos et al., 2011). This process that FOSP follow to accumulate resources, together with the limited availability of resources in the market and the high competition for them, generates a disparate ecosystem with the scale-free characteristic observed (Xu and Madey, 2004). Nevertheless, as FOSP are social products, there are ways in which project leaders can act to influence how attractive their projects are to the resources they are interested in attracting and motivating.

The findings suggest that new users, visitors and developers are likely to choose the more attractive project in comparison to the others. Yet, new projects can be better positioned, such as by choosing an appropriate license to increase their likelihood of being chosen. Project leaders should be mindful of how certain decisions affect visitors, users and developers perceptions similarly, easing their task of managing the community and software evolution.

New and undifferentiated projects stand a small chance to succeed in a populous and competitive environment, as the influence of project characteristics is relatively modest when comparing to the self-reinforcing force of attractiveness. Nevertheless, there are ways to overcome this ''liability of newness''. For example, one may reuse available open source code and create derivative projects (forking), or opensource a proprietary software with an established user-base and good reputation, carefully choosing which characteristics to give to the new project (e.g., avoiding being listed in the application domain ''others'' or choosing a license strategy that differentiates the project from its competitors).

Some of the project conditions affect attractiveness and work activities in opposite directions, such as the GPL license and the dual-licensing strategy. This creates opportunities for the project manager to operate strategically. For example, one may decide to have dual-licensing first, focusing on recruiting resources for the new project to build critical mass, and later on change to non-restrictive licenses, focusing on the generation of work activities, which would improve attractiveness indirectly through other means such as software maintenance. Additionally, project managers should invest higher amounts of their private resources in the initial stages of a software release, as ''young'' applications (up to alpha) tend not to attract resources from the community. This means that a higher effort on work activities and other indirect contributions such as running ads on key channels is required to overcome the challenge to emerge, until the thresholds of critical mass and ''mature application'' are reached.

On a theoretical side, our findings suggest that variables such as website hits, downloads and members should be treated as causes of each other or as final dependent variables on studies of FOSP *only* in restricted and well justified situations. We have shown that these variables are highly correlated not because they cause each other, but due to the existence of causes

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

which lead to their occurrence. One decides to visit a website based on reasons similar to those one considers to download and use, or even to become a member and contribute to open source software. The process may occur in sequence, from visiting to using to developing, but it is people's perception of software value (i.e., how attractive they perceive the project to be) that influences their moving from one step to another (von Krogh and Spaeth, 2007). It is the perception of the project at consideration that impacts people's behaviour. Moreover, the presence of visitors, users and developers alone, or in combination, cannot improve a project. These agents are sources of improvements, but their presence is not improvement and could only sustain a project attractiveness in satisfactory levels temporarily. Therefore, to treat them as independent variables (or mediators) is vital to accurately represent how user-driven innovations can be stimulated by sponsors and social planners in general (Arakji and Lang, 2007).

Our findings shed new light into a dilemma in the Information Systems literature, where one stream claims that the quantity of people increases diversity, which fosters problem-solving and innovation (Raymond, 1999; West and O'Mahony, 2005), whereas the other argues via empirical evidence that the relationships between the number of developers and software quality and project effectiveness and efficiency are not significant (Balijepally et al., 2009; Stewart and Gosain, 2006). Without claiming to close the debate, our results can conciliate these two streams by presenting effects and side-effects of increasing the number of contributors. More people tend to locate and fix more bugs, request and develop more features, creating an environment, where innovations are more likely to occur and higher-quality software generated. However, more people reduce the likelihood to solve an increasing number of bugs reported and features requested at the same time that more time becomes necessary to solve tasks, possibly due to coordination difficulties (Comino et al., 2007). Therefore, with positive and negative effects of an increasing number of contributors, both streams raise valid points and should not be seen as contradictory.

On a related note to the number of contributors in projects, our research provides a new perspective to the free-riding "problem", which is commonly reported as something that threatens the open source model (Baldwin and Clark, 2006). The traditional idea is that private agents are not expected to invest enough of their resources to produce a public good, they would rather free-ride. However, as we have shown, users and visitors, which would qualify as free-riders in the traditional sense, are actually contributing to the project as well, helping it build a critical mass and raising developers' intention to contribute. So, there is nothing open source project creators and managers should do about preventing free-riders, for besides contributing indirectly or unintentionally, they have chosen not to go to the competition and, thereby, improve project's attractiveness instead of competition's. There is no pressing need to worry about attracting free-riders, said to be reducers of the probability of success (Bessen, 2005), as FOSP can actually benefit from them in various ways. Johnson's prediction that "when more individuals are present, the incentive to free-ride is raised" (2002, p. 644), so that contributions become less likely to occur, has no support from our perspective.

Moreover, frequently, members of FOSP communities are developers and users of the software, which broadens their perspectives on software quality to contain technical, functional, and business domain dimensions. This creates an environment, where many are likely to contribute because one's contribution as a developer benefits oneself as a user. That is a recipe for success, giving these communities an advantage over software produced by an organization, where developers and users are independent entities, requiring extra effort to align needs and priorities. Accordingly, more available resources, of any kind, should indeed create an environment favourable to software quality and project success in open source. From this perspective, there is no need to manage what type of resources a project is attracting, as one type helps bring the others, and the various types of contributions come for similar reasons.

Having established the need for organizations involved in FOSP to improve their projects' attractiveness and visibility along with prior research (e.g., Agerfalk and Fitzgerald, 2008), this study is informative for the strategic software development practice. Our study can guide organizations on matters to be faced when opensourcing software, or deciding which one to sponsor, by providing insights on how attractive a software would be, and how to influence it (e.g., through the selection of an specific combination of licenses or spending more on advertising towards the beginning). Specifically, organizations may use the results strategically: (1) to identify among their software, according to application domain and type of user, the ones more likely to succeed if opensourced; (2) as a guide to design and position a project more effectively, managing its attractiveness to attain desired goals; (3) to help decide on when to release source code (stage of development), adding objectivity to the subjective advices previously discussed that software should be released in "later stages" (Johnson, 2002; Raymond, 1999); (4) to plan on what to expect from the community as the software evolves, managing better the evolution process by adjusting coordination methods and marketing efforts accordingly; and (5) and to judge which project is preferable and more appealing to developers, visitors and users, both to adopt and get involved, and thus strategically choose, where to place sponsorship resources.

Finally, as the open source model of user-driven innovation resembles the knowledge production in science and has been adopted in other fields (von Krogh and Spaeth, 2007), the research model proposed here can be adapted to help us understand other public projects of collective production as well, especially those that fit into the category of open innovation. One particular evident case outside the software industry is the production and improvement of knowledge that takes place in the public encyclopedia Wikipedia.org. By analogy, we can associate a page or article in the encyclopedia with an open source software project, and say that articles have readers (visitors), content users (e.g., those who cite), and writers (developers). As far as the distribution of these resources to articles, we expect it to be highly skewed as well, towards the popular culture and topics in fashion, for example (contextual factors). Accordingly, we can begin to construct a similar model, considering articles' set of conditions that influence resources behaviour (e.g., stage of development, language, type of content, etc.), gathering different types of contributions to articles such as orthographic corrections, addition of paragraphs,

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

usage by citing the article in academic papers, so on and so forth. Probably, this model of collective production would generate new insights on how to design articles that are more likely to attract an independent and voluntary community of contributors around them, sustaining knowledge production and dissemination by and to public. A similar theoretical exercise could be performed to generate models for blogs, social networks' profiles, open innovation problem solving, email list threads, etc. In conjunction, these various models would create a body of knowledge useful for organizations and individuals to operate strategically and so more effectively in their Internet-based strategic endeavours of communicating and collaborating with peers and customers.

7. Limitations

Research limitations can be divided into internal and external. Internal limitations are related to how the data was collected and the analysis performed. First, on capturing the effects of license, we could not classify all licenses available to FOSP according to their restrictiveness. We ignored licenses not classified by Lerner and Tirole (2005) and thus lost their effects. Second, although we collected data over time, it was analysed in the structural model using a cross-sectional approach. In doing so, we were not able to control for auto-correlations. However, this limitation is a cost of using a public secondary data in need of validation. To some extent, this limitation was addressed by our mediational model, which provided further support for the claims that FOSP characteristics and the number of contributors influence together how a project grows, changes, and gains momentum over time.

Amid the external limitations, a variety of potentially important variables were omitted from the analysis. There are other candidate explanations for attractiveness that were not included in our model. Namely, we identified: (a) members' technical knowledge, directly affecting the types of tasks generated and addressed; (b) level of community trust on the project and its sponsors, affecting people's willingness to contribute (Agerfalk and Fitzgerald, 2008); (c) existence of sponsored developers, who would keep a project active on a regular basis, having a formal commitment to address tasks; (d) the possibility of an open source project be included in official GNU/Linux distributions, influencing directly its accessibility; (e) the effects of the programming language adopted as the availability of developers depends on that (Stewart et al., 2006); and (f) the role of usability and source code metrics on project attractiveness (Rajanen and Iivari, 2010; Meirelles et al., 2010). This list of causes is exemplary and many other possible influencers exist, but given that empirical research is constrained by mundane restrictions, we believe that our model is useful for it communicates the message and provides ground for further verifications and refinements.

Additionally, we believe that the "self" effect of attractiveness is a proxy for many mediators, which could not be measured. For example, when users spend their money and time promoting the project, the actual chain of causality is from attractiveness, to finding a user, to receiving a contribution (money and time) and, finally, to attractiveness. The attractiveness to attractiveness proposition is simply a shortcut, useful for empirically verifying the theoretical argument and for prediction purposes. Future research should measure these unobserved mediators.

Moreover, there is the fact that a project may change its characteristics over time, which our empirical model assumed to be stable. However, we believe that in such case, the theoretical model we laid out would still be valid. When a project changes its conditions, for example, its type of license, we expect that the change would trigger an effect on the project attractiveness and activities in the terms we have already discussed. Previous empirical research supports this claim, showing that the decision to change license alters project attractiveness (Santos et al., 2011).

As a final limitation, we admit that "a volunteer-based community [...] may not behave strategically", as Jaisingh et al. (2008, p. 260) stated. In that case, our study would lose some of its explanatory power in exchange to strengthening its prescriptive nature. Nevertheless, with higher prescriptions, this study demonstrates how opportunities to operate strategically in the (open source) software market exist and may be used to maximize social welfare (Jaisingh et al., 2008). For the latter, governments may reduce software development costs by effectively opensourcing their applications and choosing the most attractive ones to adopt, and countries with less capability to develop software would have easier access to the knowledge embedded in the source code and thus be better equipped to provide services to their populations by relying on what others have made available on the Web.

8. Future work

The best way to address research limitations is with follow-up studies. A wide variety of studies can be derived from our results and conclusions, covering topics related to both content and method. On the content side, as the phenomenon of dual-licensing was just recently identified and discussed (Jiang and Sarkar, 2009; Watson et al., 2008), and, to the best of our knowledge, this is the first study to empirically assess the impacts of this choice, replications would be beneficial. Moreover, Agerfalk and Fitzgerald (2008) pointed out that the recruitment of developers from an open source project by sponsors could erode the "unknown" aspect of the project, affecting trust levels and innovation rates. Their proposition relates to the limitation of not statistically controlling for sponsored-contributors and reinforces the need to add it in future studies.

Additionally, we encourage the development of competing models of attractiveness to be compared with this one, which is exploratory in nature. For example, we have stated that project activities, as well as users and developers, are consequences of attractiveness and decided to group visitors, users and developers because they are the contributors. Thus,

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

a competing model with all of them as indicators of one latent construct should also be evaluated. However, we have performed preliminary analysis and found that the addition of the variables related to project activities generates a principal component analysis solution with two factors (1-Eigenvalue = 3.38, variance explained = 48%, 2-Eigenvalue = 1.08, variance = 15%). The complete solution has seven components and thus supports our original proposition that only the sources of contributions, and not their contributions, should be grouped together.

Similar to prior research in code reuse (Maillart et al., 2008), our paper addresses attractiveness from the preferential attachment perspective. But this does not preclude other plausible mechanisms including the 'like attracts like', especially during the earlier stage of project formation. It is likely that the like attracts like mechanism (as predicted in script theory and social status hierarchy) presents a critical resource at the initial stage, in that a small number of highly resourceful developers of similar status will work together initially (Kuk, 2006). Once the critical mass is reached, the rich-gets-richer effect takes over (Lee et al., 2006) as the emergent networks characterized by peripheral contribution will be less costly to maintain (Hu and Wang, 2009). Future research can seek to map out these longitudinal changes of what explains the preferable projects.

Furthermore, as previously shown, the concept of attractiveness can be adapted and used in other fields of research, being useful to any user-centric effort or collaborative work performed on the Web (e.g., see Comino et al., 2007; Wagner and Majchrzak, 2007). In addition, the relative strategic value of opensourcing software, when compared to the traditional outsourcing and insourcing approaches (Qu et al., 2010), should also be studied. Finally, the theoretical framework developed in this paper can be extended by incorporating results of previous research, such as the explanation of what determines license choices (Sen et al., 2008). In doing that, we can visualize a more complete model that explains what determines the license choices, which impacts user adoption and contributor recruitment, project activities, and, ultimately, software maintenance and improvement, all of that under the perspective of FOSP attractiveness.

On the method side, variables' effect-sizes were not calculated by us. Accordingly, we do not know how strongly a specific project activity variable influences project attractiveness, for example. Also, the results reported here could be further evaluated in an attempt to predict projects' future attractiveness at various distances in time (lags). Thus, without future research, we will not know for how long a state of attractiveness lasts or is capable of sustaining recruitment, participation and engagement.

9. Conclusions

Given the increasingly common model of developing software by dissolving organizational boundaries and decentralizing the work activities to interact with globally distributed workers and users, this paper started by identifying and describing the distributional characteristics of FOSP key-resources in order to build a theoretical explanation for their behaviour. That is, we relied on the existent literature that demonstrated that most users and developers of open source are concentrated in a few projects for reasons yet unidentified but that conform with the preferential attachment phenomenon. We proposed that FOSP attractiveness is the fundamental reason for this observed behaviour, and provided an explanation for how the majority of visitors, users and developers come to select only a few projects to adopt and contribute, ignoring the rest (Xu and Madey, 2004).

Additionally, we stated that the notion of contributor and contribution so far adopted in the FOSP literature has been mostly restricted to source code developers and, only rarely, users or visitors. In doing so, many types of contributions have been neglected along with their role in project improvement and software maintenance via project activities and other less direct, or even unintended, modes of contributing.

After developing the theoretical framework, this paper empirically analysed data on thousands of projects from different sources to unfold patterns in their internal activities, which are consistent with the framework developed, according to complementary and independent statistical techniques. In summary, the results inform us about FOSP on a variety of areas. In the model background, there is a pool of FOSP created as a result of opensourcing initiatives, sometimes utilized "as a marketing technique" (Jiang and Sarkar, 2009, p. 208); and there is a community interested in using, studying and contributing to these projects. Thus, our concerns were to explore what drives people to specific projects, or what types of projects are more attractive to people, understanding how that impacts project activities and, consequently, attractiveness, cyclically and continuously.

We found out that attractiveness may indeed be a strong driving force of FOSP dynamics, working as magnetic core that influences how several relevant variables are related to each other in a systemic, reinforcing, way. In a nutshell, the conclusion is that the influx of resources and contributions in FOSP depends on project attractiveness, which is a product of contributions of many kinds that come more frequently and with higher intensity depending on projects' set of conditions. Accordingly, the theme of highest value to organizations interested in releasing or sponsoring open source software as a strategic choice is how to set up and manage a project to influence its attractiveness, selecting, designing and coordinating it to that end.

Acknowledgments

This research was funded by FAPESP (www.fapesp.br, process: 2009/02046-2), and the Horizon Digital Economy Research Institute at the University of Nottingham (http://www.horizon.ac.uk/). The sponsors had no influence on the decision to publish or in the content of the paper.

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

References

Agerfalk, P.J., Fitzgerald, B., 2008. Outsourcing to an unknown workforce: exploring opensourcing as a global sourcing strategy. MIS Quarterly 32, 385–409.
Anderson, J.C., Gerbing, D.W., 1998. Structural equation modeling in practice: a review and recommended two-step approach. Psychological Bulletin 103, 411–423.
Arakji, R., Lang, K., 2007. Digital consumer networks and producer–consumer collaboration: innovation and product development in the video game industry. Journal of Management Information Systems 24 (2), 195–219.
Bagozzi, Richard P., Dholakia, Utpal M., 2006. Open source software user communities: a study of participation in Linux user groups. Management Science 52 (7), 1099–1115.
Baldwin, C.Y., Clark, K.B., 2006. The architecture of participation: does code architecture mitigate free riding in the open source development model? Management Science 52 (7), 1116–1127.
Balijepally, V., Mahapatra, R., Nerur, S., Price, K.H., 2009. Are two heads better than one for software development? The productivity paradox of pair programming. MIS Quarterly 33, 91–118.
Barabási, A.L., 2005. Network theory – the emergence of creative enterprise. Science 308, 639.
Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.
Baron, R.M., Kenny, D.A., 1986. The moderator–mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations. Journal of Personality and Social Psychology 51, 1173–1182.
Bentler, P.M., 1989. EQS Structural Equations Program Manual. BMDP Statistical Software Inc., Los Angeles, CA.
Bessen, J.E., 2005. Open Source Software: Free Provision Of Complex Public Goods. Technical Report. <http://ssrn.com/abstract=588763>.
Bevan, N., 2006. Practical issues in usability measurement. Interactions 13 (6), 42–43.
Byrne, B., 2006. Structural Equation Modeling With EQS: Basic Concepts, Applications, and Programming, Multivariate Applications Series. Lawrence Erlbaum Associates.
Chin, W., 1998. Issues and opinion on structural equation modeling. MIS Quarterly 22.
Clauset, A., Shalizi, C.R., Newman, M.E.J., 2009. Power-law distributions in empirical data. SIAM Review 51, 661–703.
Colazo, J., Fang, Y., 2009. The impact of license choice on open source software development activities. Journal of the American Society for Information Science and Technology 60 (5), 997–1011.
Comino, S., Manenti, F.M., Parisi, M.L., 2007. From planning to mature: on the success of open source projects. Research Policy 36 (10), 1575–1586.
Crowston, K., Howison, J., 2006. Hierarchy and centralization in free and open source software team communications. Knowledge, Technology, and Policy 18, 65–85.
Crowston, K., Scozzi, B., 2002. Open source software projects as virtual organizations: competency rallying for software development. IEE Proceedings Software Engineering 149, 3–17.
Crowston, K., Annabi, H., Howison, J., 2003. Defining open source software project success. In: 24th International Conference on Information Systems (ICIS), Seattle, WA.
Crowston, K., Annabi, H., Howison, J., Masango, C., 2004. Effective Work Practices for Software Engineering: Free/Libre Open Source Software Development. WISER Workshop on Interdisciplinary Software Engineering Research, SIGSOFT, Newport Beach, CA.
Crowston, K., Annabi, H., Howison, J., Masango, C., 2005. Towards a portfolio of FLOSS project success measures. In: 26th International Conference on Software Engineering, Edinburgh, UK.
Dow, K.E., Jackson, C., Wong, J., Leitch, R.A., 2008. A comparison of structural equation modeling approaches: the case of user acceptance of information systems. Journal of Computer Information Systems 48, 106–114.
Fang, Y., Neufeld, D., 2009. Understanding sustained participation in open source software projects. Journal of Management Information Systems 24 (4), 9–50.
Fershtman, C., Gandal, N., 2007. Open source software: motivation and restrictive licensing. International Economics and Economic Policy 4, 209–225.
Fershtman, C., Gandal, N., 2011. Direct and indirect knowledge spillovers: the "social network" of open-source projects. The RAND Journal of Economics 42, 70–91.
Fitzgerald, B., 2006. The transformation of open source software. MIS Quarterly 30, 587–598.
Grewal, R., Lilien, G.L., Mallapragada, G., 2006. Location, location, location: how network embeddedness affects project success in open source systems. Management Science 52, 1043–1056.
Guimera, R., Uzzi, B., Spiro, J., Amaral, L.A., 2005. Team assembly mechanisms determine collaboration network structure and team performance. Science 308 (5722), 697–702.
Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L., 2006. Multivariate Data Analysis. Pearson Education Inc., Upper Saddle River, NJ.
Herbsleb, J., Mockus, A., 2003. An empirical study of speed and communication in globally distributed software development. IEEE Transactions on Software Engineering 29, 481–494.
Hu, H., Wang, X., 2009. Disassortative mixing in online social networks. A letter. Journal Exploring the Frontiers of Physics 86.
Jaisingh, J., See-To, E., Tam, K., 2008. The impact of open source software on the strategic choices of firms developing proprietary software. Journal of Management Information Systems 25 (3), 241–275.
Jiang, Z., Sarkar, S., 2009. Speed matters: the role of free software offer in software diffusion. Journal of Management Information Systems 26 (3), 207–239.
Johnson, J.P., 2002. Open source software: private provision of a public good. Journal of Economics & Management Strategy 11, 637–662.
Judd, C.M., Kenny, D.A., 1981. Process analysis: estimating mediation in treatment evaluations. Evaluation Review 5, 602–619.
Kaplan, D., 2008. Structural Equation Modeling: Foundations and Extensions. Sage, London.
Keller, E.F., 2005. Revisiting "scale-free" networks. BioEssays 27, 1060–1068.
Kelloway, E.K., 1998. Using Lisrel for Structural Equation Modeling. International Educational and Professional Publisher, SAGE Publications, CA.
Kline, R.B., 1998. Principles and Practice of Structural Equation Modeling. The Guilford Press, New York, NY.
Koch, S., 2004. Profiling an open source project ecology and its programmers. Electronics Markets, Special Section: Open Source Software, 14.
Krishnamurthy, S., 2002. Cave or Community? An Empirical Examination of 100 Mature Open Source Projects. First Monday, 7 (6). <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/960/881>.
Kuk, G., 2006. Strategic interaction and knowledge sharing in the KDE developer mailing list. Management Science 52, 1031–1042.
Lee, S.H., Kim, P.J., Jeong, H., 2006. Statistical properties of sampled networks. Physical Review Letters 73.
Lerner, J., Tirole, J., 2002. Some simple economics of open source. Journal of Industrial Economics 50 (2), 197–234.
Lerner, J., Tirole, J., 2005. The scope of open source licensing. Journal of Law, Economics and Organization 21, 20–56.
Long, J., 2006. Understanding the role of core developers in open source software development. Journal of Information, Information Technology, and Organizations 1.
Maillart, T., Sornette, D., Spaeth, S., von Krogh, G., 2008. Empirical tests of Zipf's law mechanism in open source Linux distribution. Physical Review Letters 101.
Maitland, S.B., Dixon, R.A., Hultsch, D.F., Hertzog, C., 2001. Well-being as a moving target: measurement equivalence of the Bradburn affect balance scale. Journal of Gerontology: Psychological Sciences 56 (2), 69–77.
Mardia, K., Kent, J., Bibby, J., 1980. Multivariate Analysis. Academic Press, NewYork.
Meirelles, P., Santos Jr., C., Terceiro, A., Miranda, J., Chavez, C., Kon, F., 2010. A study of the relationships between source code metrics and attractiveness in free software projects. In: Brazilian Symposium on Software Engineering (SBES), Salvador, Brazil.

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004

Mockus, A., Fielding, R.T., Herbsleb, J., 2000. A case study of open source software development: the Apache server. In: Proceedings of the 22nd International Conference on Software Engineering.
Muffato, M., 2006. Open Source: A Multidisciplinary Approach. Imperial College Press, London, UK.
Mulaik, S.A., 2005. Parsimony/Occham's Razor. Encyclopedia of Statistics in Behavioral Science.
Newman, M.E.J., 2002. Assortative mixing in networks. Physical Review Letters 89, 34–234.
Nonnecke, B., Preece, J., 2003. Silent participants: getting to know lurkers better. In: Lueg, C., Fisher, D. (Eds.), From Usenet to CoWebs: Interacting with Social Information Spaces. Springer Verlag.
Nonneke, B., Preece., J., 2000. Lurker demographics: counting the silent. In: Proc. SIGCHI Conf. Human Factors Comput. Systems (ACM), New York, pp. 73–80.
O'Mahony, S., 2007. The governance of open source initiatives: what does it mean to be community managed? Journal of Management & Governance 11, 139–150.
Papadakis, E.N., Tsionas, E.G., 2010. Multivariate Pareto distributions: inference and financial applications. Communications in Statistics – Theory and Methods 39 (6), 1013–1025.
Peterson, R.A., 1994. A meta-analysis of Cronbach's coefficient alpha. Journal of Consumer Research 21, 381–391.
Preacher, K.J., Hayes, A.F., 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavior Research Methods 40, 879–891.
Price, D.J.de S., 1976. A general theory of bibliometric and other cumulative advantage processes. Journal of the American Society for Information Science 27, 292–306.
Qu, W.G., Oh, W., Pinsonneault, A., 2010. The strategic value of IT in sourcing: an IT-enabled business process perspective. Journal of Strategic Information Systems 19 (2), 96–108.
Raja, U., Tretter, M., 2006. Investigating open source project success: a data mining approach to model formulation, validation and testing. In: Proceedings of SUGI 31, San Francisco, California.
Rajanen, M., Iivari, N., 2010. Traditional usability costs and benefits - fitting them into open source software development. In: 18th European Conference on Information Systems (ECIS), Pretoria, SA.
Raymond, E.S., 1999. The Cathedral and the Bazaar. O'Reilly, Sebastopol, CA.
Roberts, Jeffrey A., Hann, Il-Horn, Slaughter, Sandra A., 2006. Understanding the motivations, participation, and performance of open source software developers: a longitudinal study of the apache projects. Management Science 52, 984–999.
Rutner, P.S., Hardgrave, B.C., McKnight, D.H., 2008. Emotional dissonance and the information technology professional. MIS Quarterly 32, 635–652.
Santos Jr., C., Cavalca, M., Kon, F., Singer, J., Ritter, V., Regina, D., Tsujimoto, T., 2011. Intellectual Property Policy and Attractiveness: A Longitudinal Study of Free and Open Source Software Projects. ACM Computer Supported Cooperative Work (CSCW), Hangzhou, China.
Santos Jr., C., 2008. Understanding partnerships between corporations and the open source community: a research gap. IEEE Software 25.
Sauer, R.M., 2007. Why develop open-source software? The role of non-pecuniary benefits, monetary rewards, and open-source licence type. Oxford Review of Economic Policy 23, 605–619.
Sen, R., Subramaniam, C., Nelson, M., 2008. Determinants of the choice of open source software license. Journal of Management Information Systems 25 (3), 207–239.
Shaikh, M., Cornford, T., 2003. Version Management Tools: CVS to BK in the Linux Kernel. COSPA Knowledge Base. <http://pascal.case.unibz.it/retrieve/2770/shaikhcornford.pdf>.
Sharma, S., Sugumaran, V., Rajagopalan, B., 2002. A framework for creating hybrid-open source software communities. Information Systems Journal 12, 7–25.
Simon, H.A., 1955. On a class of skew distribution functions. Biometrika 42, 425–440.
Stevens, J., 1986. Applied Multivariate Statistics for the Social Sciences. Hillsdale, NJ.
Stewart, K., Gosain, S., 2006. The impact of ideology on effectiveness in open source software development teams. MIS Quarterly 30, 291–314.
Stewart, K., Ammeter, A., Maruping, L., 2006. Impact of license choice and organizational sponsorship on success in open source software development projects. Information Systems Research 17 (2), 136–144.
von Hippel, E., 2005. Democratizing Innovation. MIT Press, Boston, MA.
von Hippel, E., von Krogh, G., 2003. Open source software and the "Private-Collective" innovation model: issues for organization science. Organization Science 14.
von Krogh, G., 2002. The communal resource and information systems. Journal of Strategic Information Systems 11 (2), 85–107.
von Krogh, G., Spaeth, S., 2007. The open source software phenomenon: characteristics that promote research. Journal of Strategic Information Systems 16, 236–253.
von Krogh, G., von Hippel, E., 2006. The promise of research on open source software. Management Science 52, 975–983.
von Krogh, G., Spaeth, S., Lakhani, K.R., 2003. Community, joining, and specialization in open source software innovation: a case study. Research Policy 32, 1217–1241.
Wagner, C., Majchrzak, A., 2007. Enabling customer-centricity using Wikis and the Wiki way. Journal of Management Information Systems 23 (3), 17–43.
Watson, R.T., Boudreau, M.-C., York, P.T., Greiner, M.E., Wynn Jr., D., 2008. The business of open source. Communications of the ACM 51, 41–46.
West, J., O'Mahony, S., 2005. Contrasting community building in sponsored and community founded open source projects. In: Proceedings of the 38th Annual Hawai International Conference on System Sciences, Waikoloa, Hawaii.
Wiggins, A., Howison, J., Crowston, K., 2009. Heartbeat: measuring active user base and potential user interest in FLOSS projects. In: Proceedings of the Fifth International Conference on Open Source Systems (OSS), pp. 94–104.
Xu, J., Madey, G., 2004. Exploration of the open source software community. In: Proceedings of North American Association for Computational Social and Organizational Science (NAACSOS), Pittsburgh, PA, USA.
Ye, Y., Kishida, K., 2003. Toward an understanding of the motivation Open Source Software developers. In: Proceedings of the 25th International Conference on Software Engineering. Portland, Oregon, pp. 419–429.
Yuan, K.-H., 2005. Fit indices versus test statistics. Multivariate Behavioral Research 40, 115–148.

Please cite this article in press as: Santos, C., et al. The attraction of contributors in free and open source software projects. J. Strateg. Inform. Syst. (2012), http://dx.doi.org/10.1016/j.jsis.2012.07.004