



European
University
Institute

DEPARTMENT
OF LAW

Accountability e Transparência

Marco Almada

Instituto Universitário Europeu



Plano geral da apresentação

1. Definições iniciais
2. Transparência e opacidade
3. Papeis da transparência
4. Relação entre transparência e *accountability*
5. Direito à explicação



Transparência

- Conceito informacional
- Ligada à abertura de um sistema para o observador externo
- Sistema de observação e conhecimento (Ananny e Crawford 2018)

Accountability

- Prestação de contas
- Em última instância, indicação de responsabilidade por atos
- Sistemas de controle de processos e resultados (Kaminski 2019b)



- Um sistema é dito *opaco* quando algo impede que ele seja percebido de forma transparente
- Opacidade não é um problema único da IA
- Opacidade pode se dar em diversos níveis (Ashby 1954, Brkan 2020):
 - Caixas brancas: temos um modelo do interior
 - Caixas pretas: conhecimento apenas empírico
 - Caixas cinzas: um pouco de cada



Burrell (2016) apresenta uma tipologia da opacidade em aprendizado de máquina:

- Opacidade resultante de fatores institucionais
- Opacidade resultante do conhecimento especializado envolvido nos sistemas
- Opacidade resultante de propriedades matemáticas e escala do sistema



- Legitimidade do sistema enquanto tomador de decisão
 - Varia com o contexto
 - Particularmente crítica em casos em que sistemas automatizados tomam decisões impactantes (Brennan-Marquez e Henderson 2019)
- Controle de qualidade das decisões
- Avaliação do nível de confiança cabível em relação a um sistema



Transparência no controle de qualidade de decisões

- Decisões automatizadas estão sujeitas a erros de diversas fontes
- Transparência em relação à existência de decisões permitiria a busca de remédios
- Transparência em relação ao processo decisório permitiria a detecção de:
 - Erros e vieses (e. g. Tischbirek 2020)
 - Injustiça epistêmica (Walmsley 2020)



Transparência e confiança em decisões

- Tratamos decisões automatizadas de forma diferente do que trataríamos decisões humanas
 - Desconfiança (Dietvorst *et al* 2014, Lee 2018)
 - Excesso de confiança (Logg 2019, Wagner 2019)
- Devemos *confiar* em sistemas automatizados?
 - Confiança como relação entre humanos (Bryson 2020)
 - Confiança em instituições



Limites da transparência

- Acompanhar não significa necessariamente ter o poder para mudar algo (Ananny e Crawford 2018, Almada 2019, Wagner 2019)
- Transparência pode afetar de forma desproporcional grupos vulneráveis (Ananny e Crawford 2018)
- Significado dos sistemas é dependente do contexto social de operação (Feenberg 1992)



- Na Europa: RGPD. Considerando 71 e Artigos 13 e 14 (Kaminski 2019a; Bibal *et al.* 2020)
- No Brasil: LGPD, art. 18 (tratamento em geral) e art. 20, § 1º (decisões automatizadas):
“O controlador deverá fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial.”
- Desafio: definição do *conteúdo* do direito



Em busca de explicações efetivas

- Explicação deve ser *funcional*, ou seja, ser orientada a realizar seus propósitos legais
 - *Accountability*
 - Avaliação de confiança
- Explicação não precisa reinventar a roda (Wischmeyer 2020)



Em busca de explicações efetivas

- Em vez de uma única explicação, devem ser adotadas explicações em múltiplos níveis (Kaminski 2019b)
 - Titulares de dados
 - Autoridades e órgãos de controle
- Explicação deve gerar informação acionável para *accountability*



Em busca de explicações efetivas

- Explicação deve levar em conta a natureza *sociotécnica* (Feenberg 1992, Ananny e Crawford 2018) dos sistemas decisores
 - Tecnologias que contribuam para transparência
 - Explicações do contexto de tomada de decisão



Em suma...

- Transparência é necessária, mas não suficiente, para assegurar *accountability*
- Opacidade não pode ser eliminada de todo, mas pode ser administrada
- Direito a explicação abrange diversas formas de comunicação a respeito dos sistemas
- Explicar um sistema envolve comunicar não só regras técnicas, mas o contexto em que o sistema opera e quais as respostas possíveis



Referências

- Almada M, ‘Human Intervention in Automated Decision-Making: Toward the Construction of Contestable Systems’, Proceedings of the 17th International Conference on Artificial Intelligence and Law (ICAAIL 2019)
- Ananny M and Crawford K, ‘Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability’: (2018) 20 New Media & Society
- Ashby WR, ‘An introduction to cybernetics’ (Chapman & Hall, 1956)
- Bainbridge L, ‘Ironies of Automation’ (1983) 19 Automatica 775
- Bibal A and others, ‘Legal Requirements on Explainability in Machine Learning’ [2020] Artificial Intelligence and Law



Referências

- Brennan-Marquez K and Henderson S, ‘Artificial Intelligence and Role-Reversible Judgment’ (2019) 109 *Journal of Criminal Law and Criminology*
- Brkan M, ‘Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond’ (2019) 27 *International Journal of Law and Information Technology*
- Brkan M and Bonnet G, ‘Legal and Technical Feasibility of the GDPR’s Quest for Explanation of Algorithmic Decisions: Of Black Boxes, White Boxes and Fata Morganas’ (2020) 11 *European Journal of Risk Regulation*
- Bryson JJ, ‘The Artificial Intelligence of the Ethics of Artificial Intelligence: An Introductory Overview for Law and Regulation’ in Markus D Dubber, Frank Pasquale and Sunit Das (eds), *The Oxford Handbook of Ethics of Artificial Intelligence* (Oxford University Press 2020)



Referências

- Burrell J, ‘How the Machine “Thinks”’: Understanding Opacity in Machine Learning Algorithms’ (2016) 3 Big Data & Society
- Dietvorst B, Simmons JP and Massey C, ‘Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err’ (2014) 116 Journal of Experimental Psychology: General
- Feenberg A, ‘Subversive Rationalization: Technology, Power, and Democracy’ (1992) 35 Inquiry
- Kaminski ME, ‘The Right to Explanation, Explained’ (2019a) 34 Berkeley Technology Law Journal
- Kaminski ME, ‘Binary Governance: Lessons from the GDPR’s Approach to Algorithmic Accountability’ (2019b) 92 Southern California Law Review



- Lee MK, ‘Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management’ (2018) *Big Data & Society*
- Logg JM, Minson JA and Moore DA, ‘Algorithm Appreciation: People Prefer Algorithmic to Human Judgment’ (2019) 151 *Organizational Behavior and Human Decision Processes*
- Obermeyer Z and others, ‘Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations’ (2019) 366 *Science* 447
- Tischbirek A, ‘Artificial Intelligence and Discrimination: Discriminating Against Discriminatory Systems’ in Thomas Wischmeyer and Timo Rademacher (eds), *Regulating Artificial Intelligence* (Springer International Publishing 2020)



Referências

- Wagner B, ‘Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems’ (2019) 11 Policy & Internet
- Walmsley J, ‘Artificial Intelligence and the Value of Transparency’ (2020) AI & Society
- Wischmeyer T, ‘Artificial Intelligence and Transparency: Opening the Black Box’ in Thomas Wischmeyer and Timo Rademacher (eds), Regulating Artificial Intelligence (Springer International Publishing 2020)