



telecom-paris.fr/ai-ethics

L'explicabilité du point de vue de la société

Winston Maxwell

Telecom Paris, Institut Polytechnique de Paris

Les lundis de l'IA et de la finance avec l'ACPR

9 novembre 2020

Cinq axes de recherche



Les biais et l'équité



IA et l'intérêt général



L'explicabilité



La responsabilité



La gouvernance et la
régulation

L'explicabilité des algorithmes

| CONTEXT FACTORS | TECHNICAL SOLUTIONS | EXPLAINABILITY PARAMETERS |
|--|--|--|
|  <p>RECIPIENT</p> <ul style="list-style-type: none">• Who is receiving the explanation,• Level of expertise,• Time available. | | |
|  <p>IMPACT</p> <ul style="list-style-type: none">• What harms possible,• Can explanation mitigate harm. | <p>POST-HOC</p> <ul style="list-style-type: none">• LIME,• Kernel-SHAP,• Saliency maps. |  <p>GLOBAL EXPLAINABILITY</p> <ul style="list-style-type: none">• User's manual,• Level of detail,• Source code,• Info on training data,• Learning algorithm,• Disclosure of biases,• Copy of training data. |
|  <p>REGULATORY</p> <ul style="list-style-type: none">• What regulatory framework,• Fundamental rights. | <p>HYBRID</p> <ul style="list-style-type: none">• Modifying objective or predictor function,• Produce fuzzy rules,• Output approaches,• Input approaches,• Genetic fuzzy logic. | |
|  <p>OPERATIONAL</p> <ul style="list-style-type: none">• Is explainability an operational imperative,• Safety certification,• Usability need. | |  <p>LOCAL EXPLAINABILITY</p> <ul style="list-style-type: none">• Counterfactual dashboards,• Saliency maps,• Level of detail,• Individual decision logs,• What information in logs, and store how long. |

Le droit à l'explication algorithmique se cherche...

| | Secteur public | Secteur privé |
|----------------|--|---|
| Les Etats-Unis | Principe constitutionnel de <i>due process</i> Facial Recognition Law de l'état de Washington (mars 2020) | FTC Act Fair Credit Reporting Act Equal Opportunity Credit Act |
| L'UE | CEDH Charte des droits fondamentaux RGPD Directive Police-Justice | RGPD Règlement « platform to business » Droit de la Concurrence |
| France | Constitution de 1958 Code des relations du public avec l'administration Loi informatique et libertés | Code de la Consommation Loi informatique et libertés Code de Commerce |

Futur règlement UE sur l'IA

La jurisprudence américaine

Loomis v. Wisconsin (Supreme Court of Wisconsin, July 13, 2016)

Houston Teachers (251 F. Supp. 3d 1168, S.D. Tex. 2017)

Le droit à l'explication relève du principe constitutionnel de *due process*.

Le poids de l'algorithme dans la prise de décision affectera le niveau d'explicabilité.

Une notice de mise en garde peut parfois suffire.

La jurisprudence européenne

Cas SyRI aux Pays Bas (Tribunal de la Haye, 5 février 2020)

CJUE La Quadrature du Net (6 octobre 2020)

L'absence d'explication peut constituer une violation de la Convention européenne des droits de l'homme.

Une recommandation algorithmique entraînant des conséquences pour l'individu doit subir un réexamen humain.

Les cinq leçons

Une explication **utile** pour celui qui le reçoit (approche règlement européen P2B)

Une décision **réplicable** (affaire *Houston Teachers*)

L'approche '**notice de mise en garde**' (affaire *Loomis*, loi *State of Washington*)

L'auditabilité par rapport aux **discriminations** (affaire *SyRI aux Pays Bas*)

Une explication qui rend le **décisionnaire humain plus intelligent** (affaire *La Quadrature du Net*)

Références



Publications

Beaudouin et al. (2020), *Flexible and context-specific AI explainability*, arXiv: 2003.07703

Beaudouin et al. (2020), *Identifying the 'right' level of explainability*, SSRN 3604924

Bertrand, Maxwell and Vamparys (2020), *Are AI-based anti-money laundering systems compatible with European fundamental rights?* SSRN 3647420

Dupont, Fliche and Yang (2020), *Governance of Artificial Intelligence in Finance*, ACPR discussion document