

ESCOLA DE COMUNICAÇÕES E ARTES – USP

Programa de Pós-Graduação em Ciências da Comunicação

Número e Algoritmo nas Interfaces Sociais da Comunicação

Prof^a Dr^a Daniela Osvald Ramos

Algoritmos e Transparência

Edison Morais
Valdir Ribeiro Jr.

600 TB/dia*

*Dados referentes a 2013.

Fonte: Blog do Facebook para programadores – <http://bit.ly/2LgCXgD>

O que é ser transparente?

"Transparency is generally considered a **means to see the truth and motives behind people's actions** (Balkin 1999) and to ensure social **accountability and trust** (Breton 2006).

"On a very basic level, **transparency allows access to more information** which can influence **power relationships** between governments and citizens, business and customers, and in our case between news outlets and audiences (Bennis 2013)."

Diakopoulos e Koliska (2016)

O que é ser transparente?

"Transparency reviews encourage the adoption of new or re-formatted informational production processes that produce information intended to fit the auspices of the review. In this way, **internal aspects of organizations are not 'made available'** but instead are re-oriented toward the production of specific forms of informational output that will externalize (**or make available**) a particular version of the **internal dynamics** of the organization"

Daniel Neyland (2007)

TABLE 1

Summary of transparency factors across four layers of algorithmic systems

Layer	Factors
Data	<ul style="list-style-type: none">• Information quality.<ul style="list-style-type: none">◦ Accuracy.◦ Uncertainty (e.g. error margins).◦ Timeliness.◦ Completeness.• Sampling method.• Definitions of variables.• Provenance (e.g. sources, public or private).• Volume of training data used in machine learning.• Assumptions of data collection.• Inclusion of personally identifiable information.
Model	<ul style="list-style-type: none">• Input variables and features.• Target variable(s) for optimization.• Feature weightings.• Name or type of model.• Software modeling tools used.• Source code or pseudo-code.• Ongoing human influence and updates.• Explicitly embedded rules (e.g. thresholds).
Inference	<ul style="list-style-type: none">• Existence and types of inferences made.• Benchmarks for accuracy.• Error analysis (including e.g. remediation standards).• Confidence values or other uncertainty information.
Interface	<ul style="list-style-type: none">• Algorithmic presence signal.• On/off.• Tweakability of inputs, weights.

Quando podemos ser transparentes?

Quando a abertura dos dados promove bem estar social. Exemplo:

- Dados agregados de segurança pública (NEV)
- Projeto serenata de amor

Quando não devemos ser transparentes?

Quando há propriedade intelectual ou de Estado envolvida, ou ao tentar burlar o sistema, os indivíduos com acesso ao código seriam beneficiados. Podendo manipular o jogo em benefício próprio.

Exemplos:

- Algoritmo do Google
- Malha fina da Receita Federal

Transparência

Transparency

- Código fonte

Accountability

- Responsabilidade, ética, prestação de contas

"Accountability must be part of the system's design from the start. Designers of such systems - and the nontechnical stakeholders who often oversee or control system design must begin with oversight and accountability in mind."

Kroll et al. (2016)

Machine Learning

A Inspeção do código-fonte não é o suficiente para a compreensão da tomada de decisão automatizada.

O código, neste caso explicita apenas o método de machine learning utilizado e não a metodologia de tomada de decisão usada.

Aleatoriedade e o Viés Cognitivo

Um algoritmo puramente randômico é um dos maiores desafios da computação.

Processos randomizados garantem isenção e equidade de análise, removem vieses humanos



"Automation introduces a surprising benefit. By limiting the role of human discretion and intuition and relying upon computer-driven decisions this process protects minorities and others weaker groups."

Zarsky, 2012

Automated Predictions: Perception, Law and Policy

COGNITIVE BIAS CODEX

What Should We Remember?

We favor simple-looking options and complete information over complex, ambiguous options

To avoid mistakes, we aim to preserve autonomy and group status, and avoid irreversible decisions

To get things done, we tend to complete things we've invested time & energy in

To stay focused, we favor the immediate, relatable thing in front of us

Need To Act Fast

To act, we must be confident we can make an impact and feel what we do is important

We Reduce Events and Lists to their Key Elements

We discard specifics to form generalities

We edit and reinforce some memories after the fact

We store memories differently based on how they were experienced

We notice things already primed in memory or repeated often

Too Much Information

We notice when something has changed

We notice flaws in others more easily than we notice flaws in ourselves

We are drawn to details that confirm our own existing beliefs

We fill in characteristics from stereotypes, generalities, and prior histories

Not Enough Meaning

We imagine things and people we're familiar with or fond of as better

We simplify probabilities and numbers to make them easier to think about

We think we know what other people are thinking

We project our current mindset and assumptions onto the past and future

Perspectivas

É possível um algoritmo ser transparente no sentido "accountability" retornando a série de pré-requisitos que está respondendo.

Para isso é necessário ser projetado **desde o começo**, por desenvolvedores éticos.

Vellido, A, Martín-Guerrero, J. and Lisboa P. 2012.
Making Machine learning models interpretable

Referências bibliográficas

- Diakopoulos, N; Koliska, M. Algorithmic Transparency in the News Media (2016). Digital Journalism.
- Neyland, D. Achieving Transparency: The Visible, Invisible and Divisible in Academic Accountability Networks (2007)
- Joshua A. Kroll , Joanna Huey , Solon Barocas , Edward W. Felten , Joel R. Reidenberg , David G. Robinson & Harlan Yu Accountable Algorithms, 165 U. Pa. L. Rev. 633 (2017).
- Vellido, A, Martín-Guerrero, J. and Lisboa P. (2012): Making Machine learning models interpretable.
- Zarsky, Tal. Automated Prediction: Perception, Law, and Policy (2012). Communications of the ACM (Volume 55, Number 9), Sept. 2012, p. 33-35.