Transparency

INFO 901 Intro to AI Ethics
March 18, 2022

Nick Diakopoulos, Northwestern University
nad@northwestern.edu | @ndiakopoulos | nickdiakopoulos.com

Goals for this Module

Goal 1. To help understand and clarify the concept of transparency both in general and as it applies to AI.

Goal 2. To develop critical perspectives and offer an opportunity to think critically and constructively about a real transparency context.

What is Transparency?

Defined as “the availability of information about an actor allowing other actors to monitor the workings or performance of this actor” (Meijer, 2014)

It is about ***information*** (both outcomes and procedures) and it is ***relational*** in that it exchanges information between actors

Why do you think transparency is relevant to AI Ethics?

How is this different an AI explanation?

Enacting Transparency for AI

Many gradations, it’s not “transparent” or “not transparent”

In general consider the *type, scope,* and *reliability* of information disclosed, as well as recipients of information, relationship b/t disclosing entity and recipient.

Distinguish *outcome transparency* (“what” the AI output) vs. *process transparency* (“how” the AI works). Relation to consequentialism and deontology.

Mechanisms for transparency: *demand-driven* (public records request), *proactive* (self disclosure) or *forced* (leaked, externally audited)

Also a communication challenge: *what* info to disclose, and *how* to disclose it

What Can be Made Transparency about AI?

Human Involvement

- Design intention/goal/purpose; involvement in design, operation, management; Intended vs out-of-scope uses; Individuals responsible for aspects of system

The Data

- Quality (accuracy, completeness, timeliness, update freq, uncertainty); bias; representativeness; provenance (motivations, funding); assumptions/limitations/transformations in preprocessing, normalizing, or cleaning data; definitions of data and quantification process; personal data used (consented);

What Can be Made Transparency about AI?

The Model / Inferences

- Features, weights, type of model; metadata about model version; threshold, heuristics, assumptions, constraints; design rationale (e.g. choice of metrics); code-level transparency; benchmarks; error analysis, confidence values or uncertainty information

Issues in Implementing Transparency

Gaming and Manipulation: Threat modeling how info might be gamed or circumvented; costs to manipulation, manipulation-resistance of data; scope

Understandability: Usable transparency; all info not for all stakeholders

Privacy: would transparency undermine PII?

Temporal Instability: Ephemerality, punctuated change, versioning,

Sociotechnical Complexity: Humans and AI are all mixed up in complex systems

Costs: $$$ to generate transparency info, documentation, benchmarks, etc.

Competitive Concerns: Undermining technical advantage; scope

Legal Context: Laws for public records; request data from corporations (credit)

Questions?

Critical Case Study Break Out

Let’s get in break out groups

Spend some time, maybe 10 minutes reading through the Facebook Community Standards Transparency Report: https://transparency.fb.com/data/community-standards-enforcement/

Discuss the report in group (take some notes to share back with the larger group)

Some Questions to Get Discussion Started

- Considering the transparency framework introduced, what questions do you have about the implementation of the FB transparency report? Are there still things that need to be made transparent?
- Do you think the transparency report is effective, why or why not?
- What kinds of critique can you apply to the implementation?
- Are there design opportunities that you think would improve the report?
- What other observations would you like to share?

Additional Reflection & Discussion

Is there anything missing from the transparency framework?

Other remaining questions you have about transparency?

Can you think of research questions stimulated from the group activity that you might pursue in your own work?

What about AI transparency in the context of research? Are you transparent enough in the AI you develop?

Thanks! Questions?

Contact

Nick Diakopoulos, nad@northwestern.edu, @ndiakopoulos
http://www.nickdiakopoulos.com/