

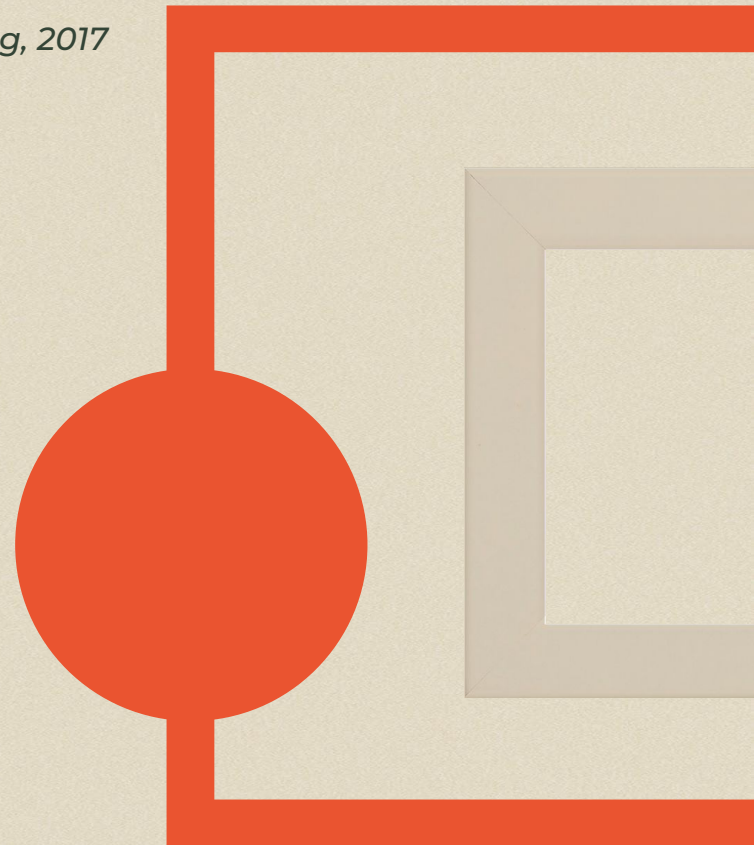
First ACL Workshop on Ethics in Natural Language Processing, 2017

Ethical by Design: Ethics Best Practices for NLP

Authors:

Jochen L. Leidner, Vassilis
Plachouras

Presented by Ebtesam Al Haque





Natural Language Processing

Teaching machines how to process and understand human language



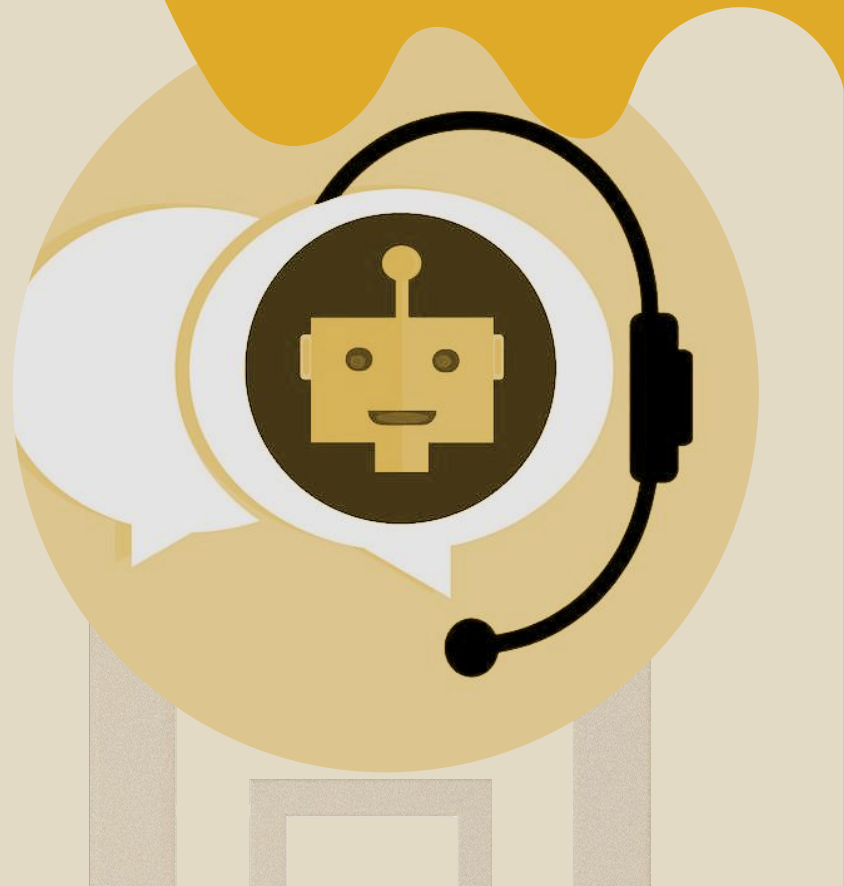
Siri



Google Translate



ChatGPT



ABSTRACT



Ethical **issues** in natural language processing



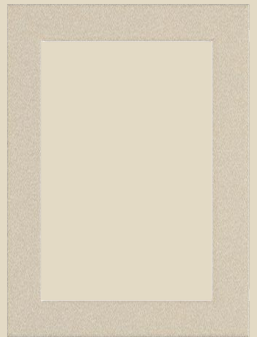
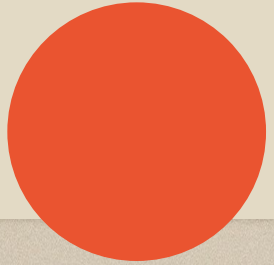
How ethical is the work, in terms of both, **process** and **outcome**?



Guidelines for best practices -
introducing *ethics review board*



First published work on NLP and ethics
from a principled process perspective



INTRODUCTION



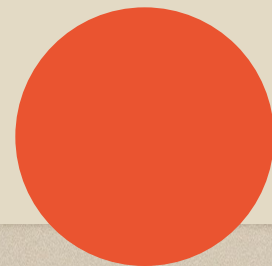
TARGET POPULATION

Software developers, practitioners

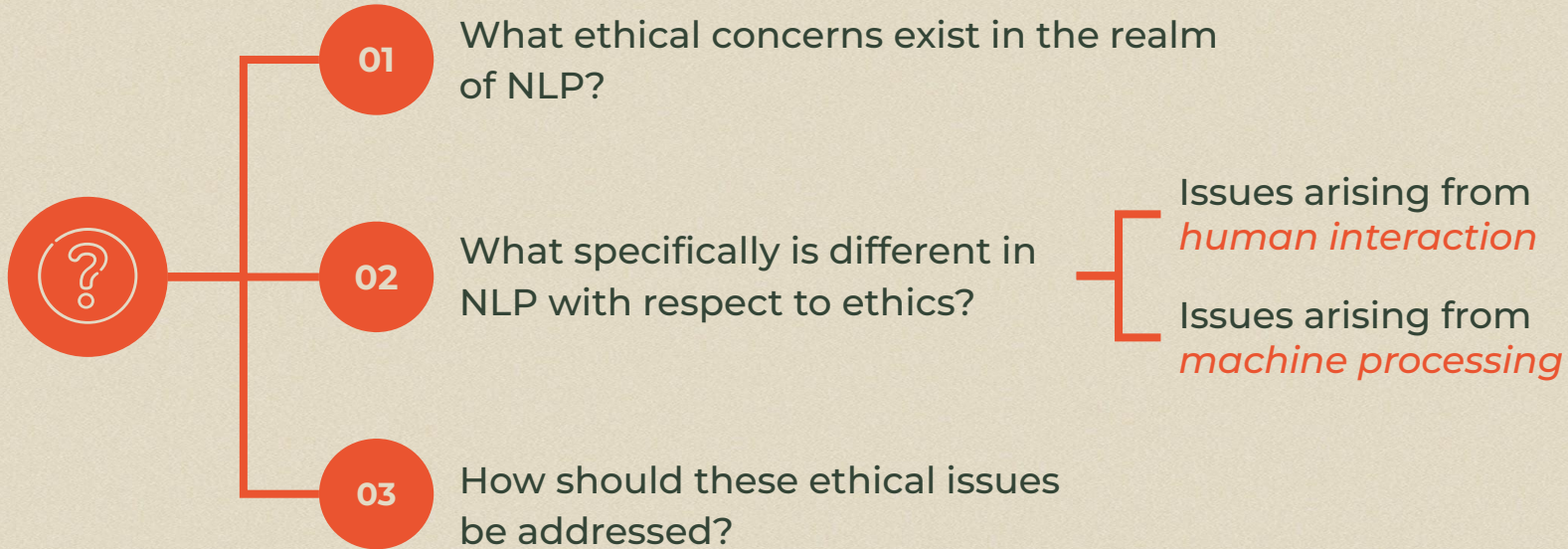


GOALS

Addressing ethical issues in NLP



RESEARCH QUESTIONS



LITERATURE REVIEW

MORAL PHILOSOPHY

Outlines three schools of moral philosophy

PROFESSIONAL ETHICS

Outlines guidelines established by various professional bodies

01

02

04

03

ML & BIG DATA ETHICS

Outlines efforts addressing issues in Big data and ML, primary focused on privacy and bias

NLP ETHICS

Outlines NLP Application and Methodology ethics

MORAL PHILOSOPHY

01

NICOMACHEAN ETHICS

Happiness is the highest attainable and ultimate goal for humans (Aristotle)

02

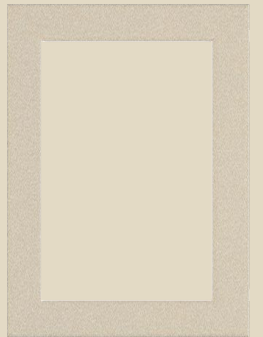
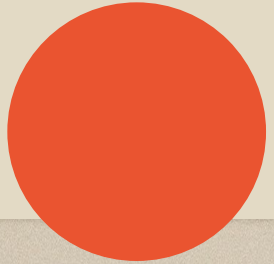
CATEGORICAL IMPERATIVE

Decision criterion to decide whether an action is morally right (Kant)

03

UTILITARIANISM

Maximize happiness for the largest number of people (Mill)



ML & BIG DATA ETHICS



INFORMATION & BIG DATA ETHICS

The *Black Box Society*, Pasquale (2015): analyzes societal impact of

- data collection
- user profiling
- data vendors and buyers

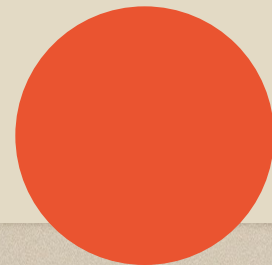


MACHINE LEARNING AND BIAS

Fairness, Accountability, and Transparency in Machine Learning (FATML) discusses *technical solutions* for

- accountability
- transparency
- fairness

in machine learning models



NLP APPLICATION ETHICS

01

SOCIAL IMPACT OF NLP

Outlines and initiates the discussion on social impact of NLP (Hovy and Spruit (2016))

02

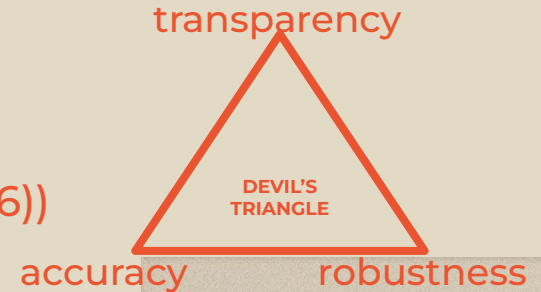
DEVIL'S TRIANGLE

Ethical considerations for chatbot detection methods (Thieltges, Schmidt and Hegelich (2016))

03

IMPACT OF NLP ON JOURNALISM

Big data and NLP creates a shift in compliance cost (Fairfield and Shtein (2014))



NLP METHODOLOGY ETHICS



CROWDSOURCING

- Ethical implications of scaling up simple tasks using anonymous human subjects
- Tasks may include
 - Image captioning
 - Question answering
 - Sentiment analysis
- Popular platform for crowdsourcing: **Amazon Mechanical Turk**
- *Encourages NLP community to seek alternative options*

(Fort, Adda and Cohen (2011))

PROFESSIONAL ETHICS

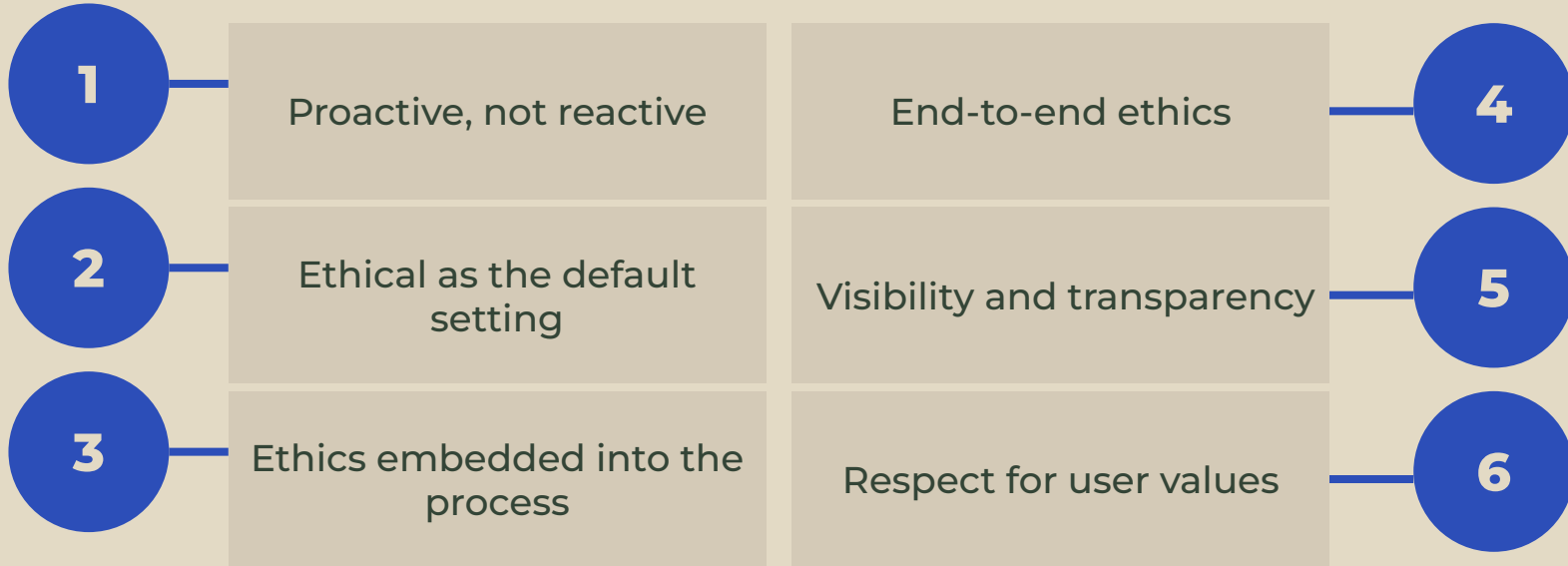
01

ACM CODE OF ETHICS

- Detailed set of 22 values, insufficient detail on certain moral boundaries, *compliance voluntary*
- 7 principles to promote *transparency* and *accountability* of algorithms: (ACM, 2017)
 1. Awareness
 2. Access and redress
 3. Accountability
 4. Explanation
 5. Data Provenance
 6. Auditability
 7. Validation and Testing

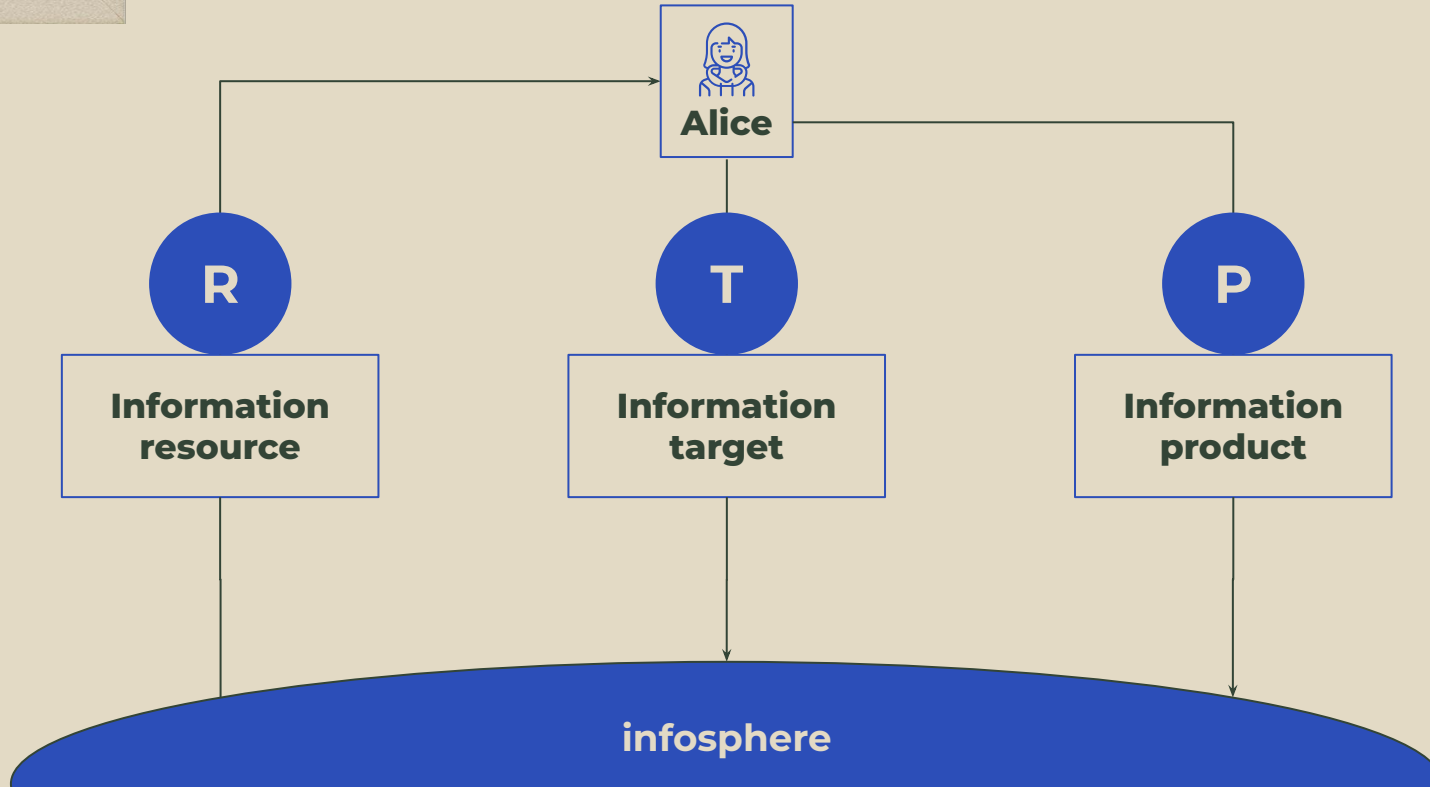
PRIVACY BY DESIGN

Principles for privacy devised by *Ann Cavoukian*, generalized



RESOURCES-PRODUCT-TARGET

Holistic “macro-ethical” model by *Floridi, 2013*





ETHICAL ISSUES OF NLP

UNETHICAL NLP APPLICATIONS

01

UNIX *spell*(1) COMMAND

Emailed words that are not found in its lexicon to its implementer, **without author's consent**, for lexicon improvements

02

AUTOMATED PSYCHOMETRICS

Predictions based on *Facebook Likes* are more accurate than those made by their friends (**Youyou, Kosinski and Stillwell (2015)**)

03

MALICIOUS CHATBOTS

Influencing a discussion or destroying it by injecting noise. *Malicious chatbots built for testing purposes must be sandboxed.*

CORE AREAS OF BIG DATA ETHICS



Davis and Patterson (2012)

T

PRIVACY

01

PERSONAL INFORMATION IN CORPORA

Some corpora contain personal information about individuals along with their identity

02

SENSITIVE PSYCHIATRIC MATERIAL

- Corpora of suicide notes compiled to study causes for terminating one's life
- building a classifier to determine how "serious" a suicide note is

03

EXCESSIVE SURVEILLANCE

Leads to self-censoring and can undermine democracy

T FAIRNESS, BIAS & DISCRIMINATION

01

INCOMPLETE TRAINING DATA

System may not exhibit similar performance across all groups of users

02

TRANSPARENCY AND ACCOUNTABILITY

Developers may not communicate all limitations of a system

ABSTRACTION & COMPARTMENTALIZATION

01

ABSTRACTION IN CROWDSOURCING

Example: Work farmed out to API can easily ignore humans behind it

02

COMPARTMENTALIZATION

Example: Building an Information extraction system without looking at the political context in which it is likely being deployed leads to ethical ignorance

T

COMPLEXITY

01

NLP API

- Big data systems today have complex pipelines that are hosted on the cloud and have *blackbox processes*
- Companies do not have much control over how their API may be used externally
- Leads to the creation of complex macrosystems beyond individual organizational boundaries

T

UNETHICAL RESEARCH METHODS

01

CONSENT

Performing experiments without briefing and getting consent from participants (and guardians, if participant is a minor)

02

WORK CONDITIONS

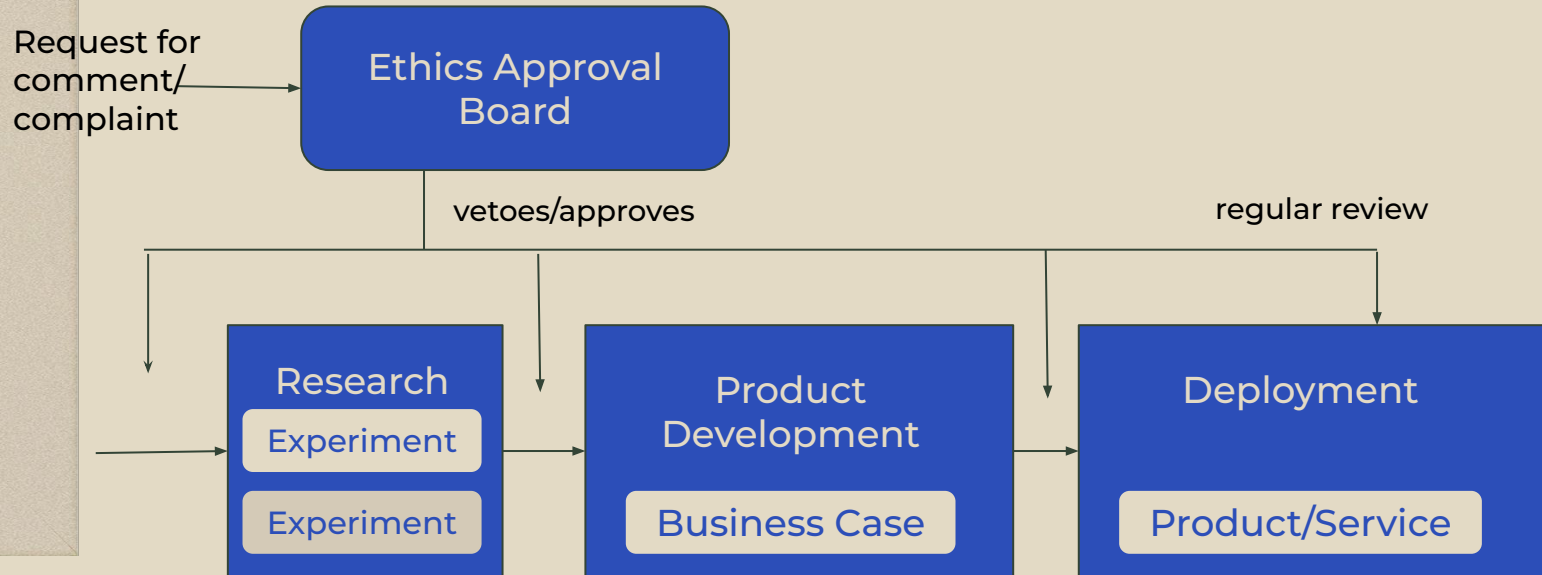
If the experiment involves efforts of helpers, they must be provided with ethical work conditions



**PROPOSED
PROCESS FOR
BEST
PRACTICES**

ETHICS REVIEW BOARD

A Process Proposal for “Ethics by Design” in an Organization.





REMEDIES FOR ETHICAL DILEMMAS

HANDLING ETHICAL DILEMMAS

DEMONSTRATION

to effect a change in society by public activism

DISCLOSURE

to document/to reveal injustice to regulators, the police,
investigative journalists
("Look what they do!", "Stop what they do!")

RESIGNATION

to distance oneself III ("I should not/cannot be part of this.")

PERSUASION

to influence in order to halt non-ethical activity ("Our
organization should not do this.")

HANDLING ETHICAL DILEMMAS

REJECTION

to distance oneself II; to deny participation; conscientious objection (“I can’t do this.”)

ESCALATION

raise with senior management/ethics boards (“You may not know what is going on here.”)

VOICING DISSENT

to distance oneself I (“This project is wrong.”)

DOCUMENTATION

ensure all the facts, plans and potential and actual issues are preserved.

DISCUSSION POINTS

1

ERB INTEGRATION

- ERB could slow down the workflow at organizations. *Can we integrate “ethics review” more efficiently? What does the tradeoff look like?*

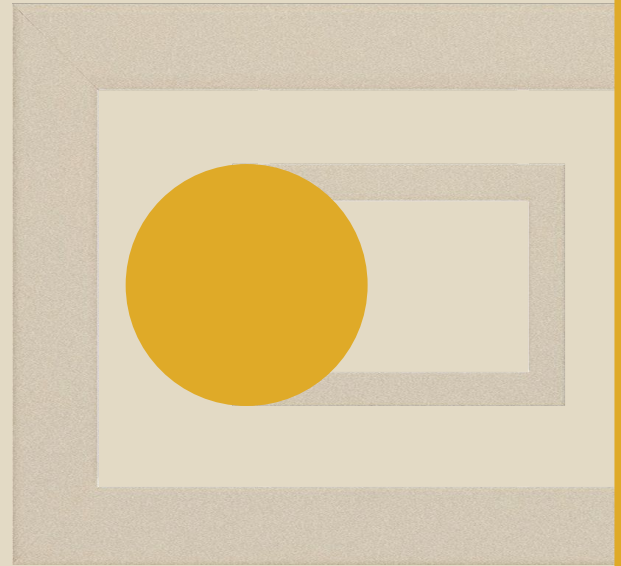
2

ERB ADOPTION

- If compliance to professional codes of conduct is voluntary, *how do we encourage practitioners to adopt ethical frameworks?*
- Would its adoption vary across *geographic regions?*

REFLECTIONS

- Given the exponential growth of NLP since 2017, we should consider doing an updated survey on its social implications. *Is it getting better? Is it getting worse?*
- Can we establish an “ethics review board” for open source projects? *How do we regulate them better?*



**THANK
YOU**

