# Critical Research and Critical Technical Practice

## A tutorial

---

Prepared and presented by

Critical Technical Practice Working Group

critical-technical-practice.github.io

# Self introduction

| | |
|---|---|
| **Mohammad Omidvar** | 5th-year PhD candidate, Reliable Systems Lab<br>Programming Analysis, Vulnerability Detection, Formal Methods |
| **Matt Canute** | Technical Lead at the Digital Democracies Institute |
| **Xueying Zhang** | 3rd-year PhD student, Tangent Lab, ixLab<br>human computer interaction, mental health technologies |
| **Weina Jin** | 5th-year PhD candidate, Medical Image Analysis Lab<br>explainable AI, medical image analysis |

## This workshop is designed to be **interactive**

self reflection activities × 1
group discussions × 1
case study × 1

how to conduct research
how to write papers

how to solve this problem
how to improve this metric
how to write papers
how to conduct peer reviews
how to design this algorithm
how to implement this system
how to conduct research
.....



**how**

how to solve this problem
how to improve this metric
how to write papers
how to conduct peer reviews
how to design this algorithm
how to implement this system
how to conduct research
.....

**why** → **how**

why should I solve this problem
why should I improve this metric

why should I design this algorithm
why should I implement this system
why do I conduct research
.....

how to solve this problem
how to improve this metric
how to write papers
how to conduct peer reviews
how to design this algorithm
how to implement this system
how to conduct research
.....

# Why asking the “why” question is important

## The official answer:

- Doing science and research is about asking “why”, to be curious, self-motivated, and keep intellectually skeptical.

## Why it matters to me:

- **Meaning** and **fulfillment.**
- **Autonomy** and **avoiding alienation**: If you don’t critically think about “why”, others will provide answers for you.

Otherwise:

- your goal can be distorted and manipulated;
- your work may lose its own **end**;
- you easily give up
- you want to jump to other problems

## why

why should I solve this problem
why should I improve this metric

why should I design this algorithm
why should I implement this system
why do I conduct research
.....

## how

how to conduct research
how to write papers
how to conduct peer reviews
how to solve this problem
how to design this algorithm
how to implement this system
.....

Self-reflection activity A1

## A1. Why do people conduct research or develop computational techniques in [fill in a field or a technique of your interest] ?

# Can the answers to the “why” question really be justified?

Referring to a commonly acceptable answer is usually the end of asking the “why” question.

But for critical research and critical technical practice, it’s the beginning.

We want to dig deeper into the why question, to continue asking “why” to the answers of the “why” question.

# Can the answers to the “why” question really be justified?

Referring to a commonly acceptable answer is usually the end of asking the “why” question.

But for critical research and critical technical practice, it’s the beginning.

We want to dig deeper into the why question, to continue asking “why” to the answers of the “why” question. → **Why?**

# Can the answers to the “why” question really be justified?

Referring to a commonly acceptable answer is usually the end of asking the “why” question.

But for critical research and critical technical practice, it’s the beginning.

We want to dig deeper into the why question, to continue asking “why” to the answers of the “why” question. → **Why?**

Referring to a commonly acceptable answer can be a way to prevent one from **questioning the status quo**.

| | Risks of asking deep "why" questions | Risks of avoiding deep "why" questions |
|---|---|---|
| Community | • Questioning and challenging the basis that underpins the status quo can be viewed as negative, disruptive, and not comfortable<br>• It may be negative for the public reputation of a research community | • The research agendas can be distorted and manipulated<br>• People follow the status quo and existing research agendas, choose low-hanging fruits and low-risk projects, stagnation of scientific discovery |
| Individual | • Research findings may be marginalized<br>• Publications and career development may face more resistance | • Suppression of one's curiosity, deep doubts, and skepticism. Their accumulation may cause mental toll and lose sense of meaning |

| | Risks of asking deep "why" questions | Benefits of asking deep "why" questions |
|---|---|---|
| Community | • Questioning and challenging the basis that underpins the status quo can be viewed as negative, disruptive, and not comfortable<br>• It may be negative for the public reputation of a research community | • Mitigating the research agendas being distorted and manipulated<br>• Keeping intellectual skepticism and questioning the status quo are the essence of science and research |
| Individual | • Research findings may be marginalized<br>• Publications and career development may face more resistance | • Being authentic to one's curiosity, deep doubts, and skepticism<br>• Gain a sense of meaning |
| | Short term risks | Long term benefits |

# Example 1. Explainable AI research

Why do people conduct research in explainable AI?

- By explaining AI predictions in human understandable ways[1], we can enhance users’ trust in the AI system

[1] Towards A Rigorous Science of Interpretable Machine Learning

Can the answers to the "why" question really be justified?

# Example 1. Explainable AI research

Why do people conduct research in explainable AI?

- By explaining AI predictions in human understandable ways[1], we can enhance users' trust in the AI system

Consequence of lack of critical reflection to the resulting explainable AI (XAI) techniques

- One dominant evaluation metric of XAI is to optimize or evaluate XAI techniques to generate as plausible explanations as possible to enhance users' trust, regardless of the underlying AI prediction correctness/quality[2]

## Systemic user manipulation and misleading

[1] Towards A Rigorous Science of Interpretable Machine Learning
[2] Why is plausibility surprisingly problematic as an XAI criterion?

Can the answers to the "why" question really be justified?

# Example 1. Explainable AI research

Why do people conduct research in explainable AI?

- By explaining AI predictions in human understandable ways[1], we can enhance users' trust in the AI system
  - How to define trust? Why providing explanations can enhance users' trust?
  - Why we need to enhance users' trust?
  - Who are "we"?
  - Enhance users' trust, for whom? For which purposes?
  - Can users really benefit from "enhancing users' trust"?
  - Does it assume a paternalistic relationship, where techniques/technical people are epistemically superior to users, and users just need to trust us and our techniques?
  - Why bear the consequences when trust is violated?

[1] Towards A Rigorous Science of Interpretable Machine Learning
[2] Why is plausibility surprisingly problematic as an XAI criterion?

Consequence of lack of critical reflection to the resulting explainable AI (XAI) techniques

- One dominant evaluation metric of XAI is to optimize or evaluate XAI techniques to generate as plausible explanations as possible to enhance users' trust, regardless of the underlying AI prediction correctness/quality[2]

## Systemic user manipulation and misleading

# Example 2. User experience design

Why do people do user experience design?

- Better understand and address users’ needs
- Improve usability and user experience
- Reduce errors and design problems
- Increase acceptance and engagement
    - Who are we designing for ?
    - What kind of understanding are we creating for our users?

Consequence of lack of critical reflection

- Prioritize experience for predefined “target users” and unintentionally exclude people outside the group.
- Prioritize a better-feeling experience over other values such as accessibility, transparency, and an accurate understanding of the technology.

# Example 3. Vulnerability detection

Why do people conduct research on vulnerability detection?

- To improve both the ability to **find** security weaknesses and to **fix** them before attackers.
  - Is a type of security vulnerability concerning in a domain?
  - Why does security issue X still happen?

Consequence of lack of critical reflection

- Devising approaches that have simpler alternatives.
- Benchmarking against projects that are not the main targets of attackers.
- Improving already-satisfying performance metrics.

# Example 4. Algorithm auditing

Why do people conduct research on auditing recommendation algorithms?

- To test for amplified mis/disinfo, filter bubbles, treating groups unfairly.
  - How does the method of bots / scraping decide which questions get asked?
  - If the algorithm is re-trained constantly, what does it mean for an audit to be "true"?

Consequence of lack of critical reflection

- Bots can like and follow easily, but they can't easily watch videos (varying dwell) the way people do, so audits focus on varying discrete signals that are easy to automate
- Repeated audits of TikTok get different results, which is reported as "poor reproducibility" [1]
- Everyday users notice harms that formal audits miss [2]

[1] Mosnar et al., Revisiting Algorithmic Audits of TikTok: Poor Reproducibility and Short-term Validity of Findings, SIGIR 2025
[2] Shen et al., Everyday Algorithm Auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors, CSCW 2021

Group discussion G1

# Can the answers to the “why” question really be justified?

Given your answer to the previous question A1

“**Why do people conduct research or develop computational techniques in** [a research field or a technique of your interest]**?**”

In a 2-3 people group, people take turns to

1. Share your answer to A1 to the group.
2. The rest people try to ask questions as **critical, pointed, and challenging** as possible to question and challenge your answer to A1.
3. Try to answer these critical questions as best as you can; it’s ok if you cannot answer them, and you can put them down in your note for you to keep thinking about them more deeply.

Group discussion G1

# Can the answers to the “why” question really be justified?

Given your answer to the previous question A1

“**Why do people conduct research or develop computational techniques in** [a research field or a technique of your interest]**?**”

Some ideas for the critical questions

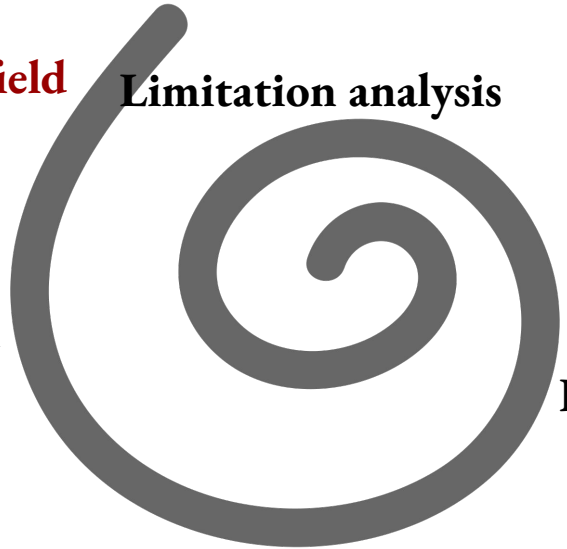
- What are the underlying assumptions? Do they hold by default?
- Whose perspective does the answer take? What would be different if taking from the potential users’ and stakeholders’ perspectives?
- Whose interests and benefits does the answer directly serve?
- Whose interests and benefits can be ignored?
- What are the potential harms and risks for the answer?
- Are there any non-technical or technical  limitations?

# Research and technical practice components

Motivation of a research field

Limitation analysis

Motivation of a research problem



Evaluation

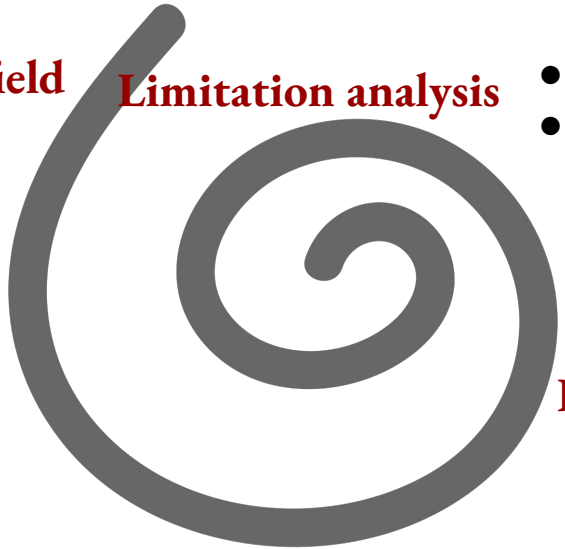
Technical design

# Common research and technical practice

## Motivation of a research field

## Limitation analysis

- Is often ignored or minimized
- Avoiding disclosing core technical limitations



## Motivation of a research problem

- Trendy topics
- Topics that are more likely to be well-received in the community
- Assumes the resulting technique is benevolent or neutral

## Evaluation

- Benchmark evaluation
- Aiming for outperforming state-of-the-art
- Cherry-picking result reporting
- Claimed benefits and functionalities may not fully be backed up with evidence

## Technical design

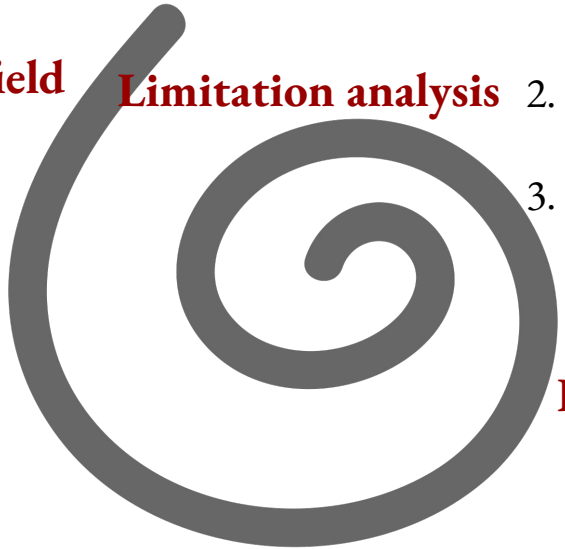
Emphasize abstracted technical approaches/theories/algorithms over concrete domain knowledge

# A more critical research and technical practice

## Motivation of a research field

## Motivation of a research problem

1. Appropriateness of problem formulation
2. Who can benefit from the development of the technique?
3. Who can potentially be put at risk due to the development of the technique?



## Limitation analysis

1. Conduct thorough assessments to understand the scope, weaknesses, harms, and failure modes of the model.
2. Disclose technique-specific limitations
3. Disclose the intrinsic limits of the modeling approach and the evaluation methods

## Technical design

Describe the domain knowledge and justify any assumptions incorporated in the model design.

## Evaluation

1. Specify the scope and level of evidence for the evaluation method
2. Avoid overclaiming and overgeneralization

## A more critical research and technical practice

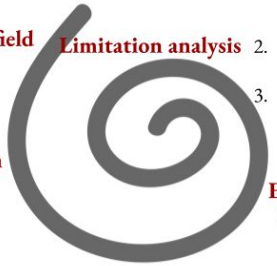
**Motivation of a research field**

**Limitation analysis**

1. Conduct thorough assessments to understand the scope, weaknesses, harms, and failure modes of the model.
2. Disclose technique-specific limitations
3. Disclose the intrinsic limits of the modeling approach and the evaluation methods

**Motivation of a research problem**

1. Appropriateness of problem formulation
2. Who can benefit from the development of the technique?
3. Who can potentially be put at risk due to the development of the technique?



**Technical design**

Describe the domain knowledge and justify any assumptions incorporated in the model design.

**Evaluation**

1. Specify the scope and level of evidence for the evaluation method
2. Avoid overclaiming and overgeneralization

26

## Common research and technical practice

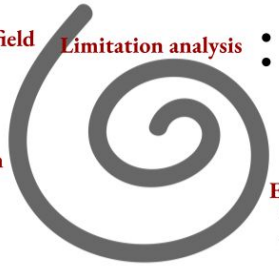
**Motivation of a research field**

**Limitation analysis**

- Is often ignored or minimized
- Avoiding disclosing core technical limitations

**Motivation of a research problem**

- Trendy topics
- Topics that are more likely to be well-received in the community
- Assumes the resulting technique is benevolent or neutral



**Technical design**

Emphasize abstracted technical approaches/theories/algorithms over concrete domain knowledge

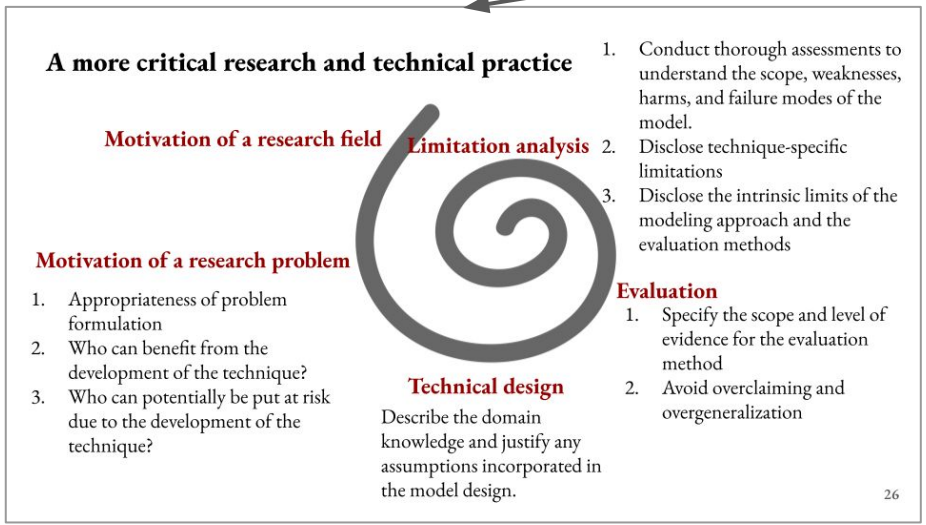
**Evaluation**

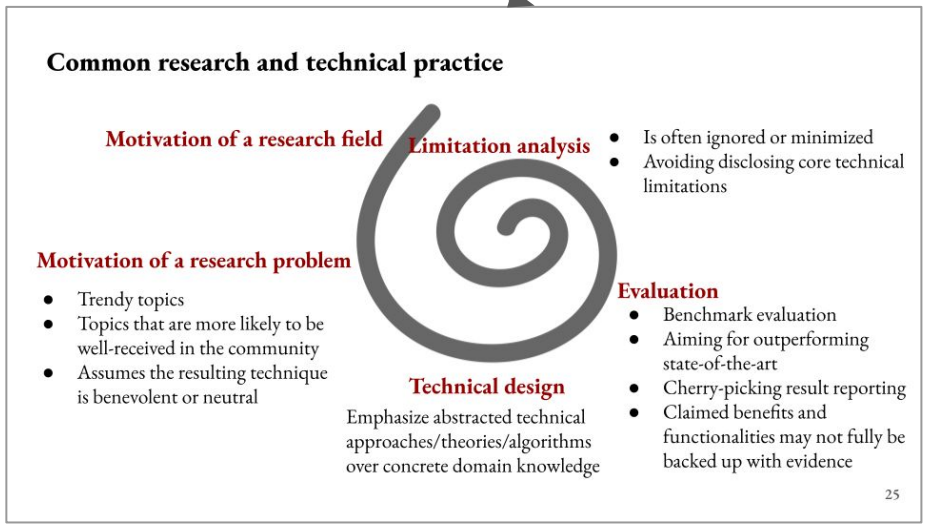
- Benchmark evaluation
- Aiming for outperforming state-of-the-art
- Cherry-picking result reporting
- Claimed benefits and functionalities may not fully be backed up with evidence

25

If models are all wrong but wrong in different ways,
then it matters to ask critically: to whom are they useful? For whose interests and benefits?

"All models are wrong, but some are useful." [1]




[1] Box, George E. P. (1976), "Science and statistics"

Case study

## What possible critiques can you think of?

### Case 1

"We show that Bayesian PMF models can be efficiently trained using Markov chain Monte Carlo methods by applying them to the Netflix dataset, which consists of over 100 million movie ratings."

### Case 2

"We have developed a new distributed agent IMPALA (Importance Weighted Actor-Learner Architecture) that not only uses resources more efficiently in single-machine training but also scales to thousands of machines without sacrificing data efficiency or resource utilisation."

Case study

## What possible critiques can you think of?

Some ideas for the critical questions and comments

1. What are the underlying assumptions? Do they hold by default?
2. What values and worldviews do the assumptions or argument imply?
3. Are there any alternative framing, assumption, values, or worldviews that you can suggest?
4. Whose perspective does the argument take? What would be different if taking from the potential users' and impacted stakeholders' perspectives?
5. By claiming the argument like this, whose interests and benefits does it directly serve?
6. Whose interests and benefits can be ignored?
7. What are the potential harms and risks of the argument?
8. Are there any non-technical or technical limitations?

### Case 1

"We show that Bayesian PMF models can be efficiently trained using Markov chain Monte Carlo methods by applying them to the Netflix dataset, which consists of over 100 million movie ratings."

### Case 2

"We have developed a new distributed agent IMPALA (Importance Weighted Actor-Learner Architecture) that not only uses resources more efficiently in single-machine training but also scales to thousands of machines without sacrificing data efficiency or resource utilisation."

# Possible critiques

Efficiency for what? for scale up? or,

for saving resources?

**Jevon's paradox**: "greater resource efficiency often leads to overall greater utilization of that resource."

"valuing efficiency facilitates and encourages the most powerful actors to scale up their computation to ever higher orders of magnitude, making their models even less accessible to those without resources to use them and decreasing the ability to compete with them."

Birhane, et al. 2022. The Values Encoded in Machine Learning Research.

**Case 1**

"We show that Bayesian PMF models can be efficiently trained using Markov chain Monte Carlo methods by applying them to the Netflix dataset, which consists of over 100 million movie ratings."

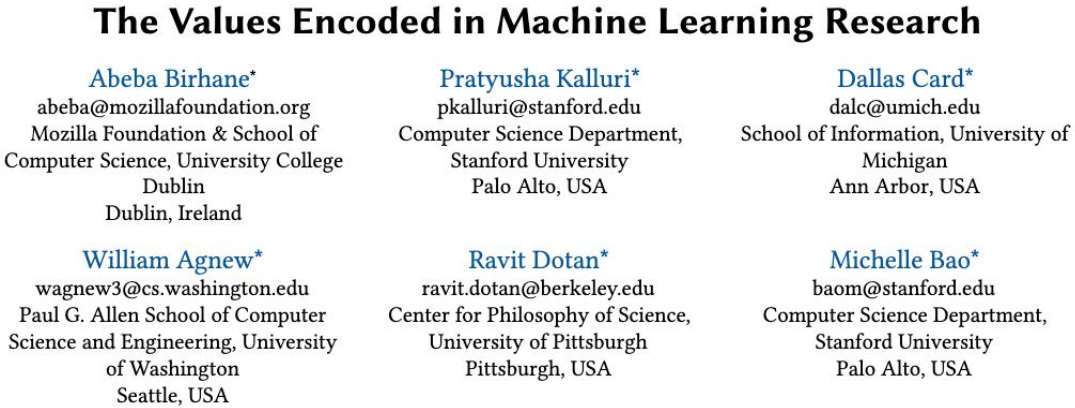
**Case 2**

"We have developed a new distributed agent IMPALA (Importance Weighted Actor-Learner Architecture) that not only uses resources more efficiently in single-machine training but also scales to thousands of machines without sacrificing data efficiency or resource utilisation."

# The cases and analysis are based on a research paper:

**The Values Encoded in Machine Learning Research.**
Birhane, et al. FAccT '22. 2022 ACM Conference on Fairness, Accountability, and Transparency

Qualitative analysis of the values encoded in
**100 highly cited NeurIPS and ICML papers** in 2008/2009 and 2018/2019



The Values Encoded in Machine Learning Research

Abeba Birhane*
abeba@mozillafoundation.org
Mozilla Foundation & School of Computer Science, University College Dublin
Dublin, Ireland

Pratyusha Kalluri*
pkalluri@stanford.edu
Computer Science Department, Stanford University
Palo Alto, USA

Dallas Card*
dalc@umich.edu
School of Information, University of Michigan
Ann Arbor, USA

William Agnew*
wagnew3@cs.washington.edu
Paul G. Allen School of Computer Science and Engineering, University of Washington
Seattle, USA

Ravit Dotan*
ravit.dotan@berkeley.edu
Center for Philosophy of Science, University of Pittsburgh
Pittsburgh, USA

Michelle Bao*
baom@stanford.edu
Computer Science Department, Stanford University
Palo Alto, USA

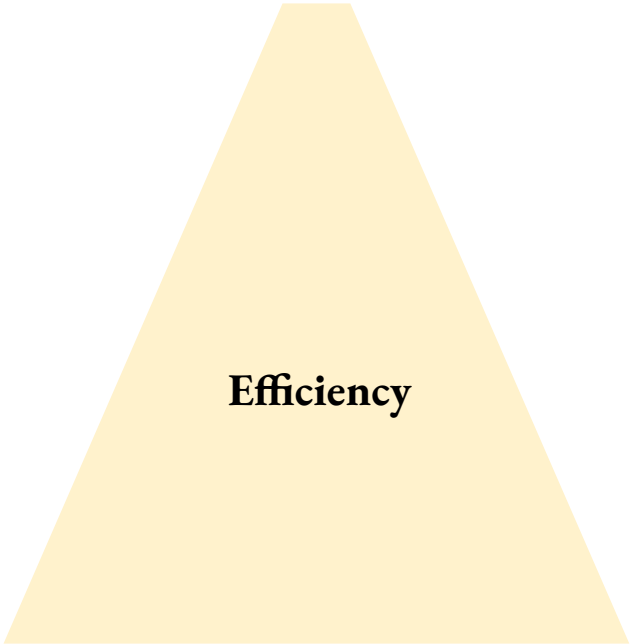
An example paper of critical research

Non-specified quotes in the case study are from this paper.

Birhane, et al. 2022. The Values Encoded in Machine Learning Research.

## Case study wrap-up

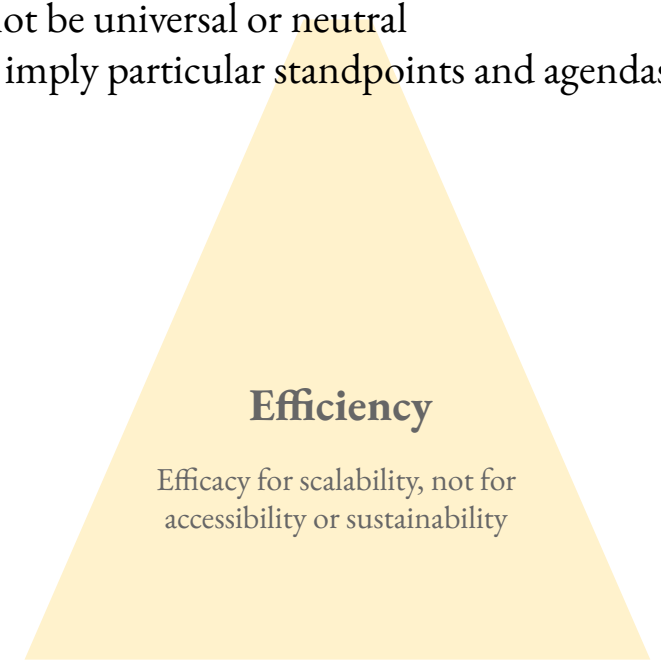
A seemingly “neutral” and benevolent value statement




Birhane, et al. 2022. The Values Encoded in Machine Learning Research.

## Case study wrap-up

A seemingly “neutral” and benevolent value statement may not be universal or neutral
It can imply particular standpoints and agendas

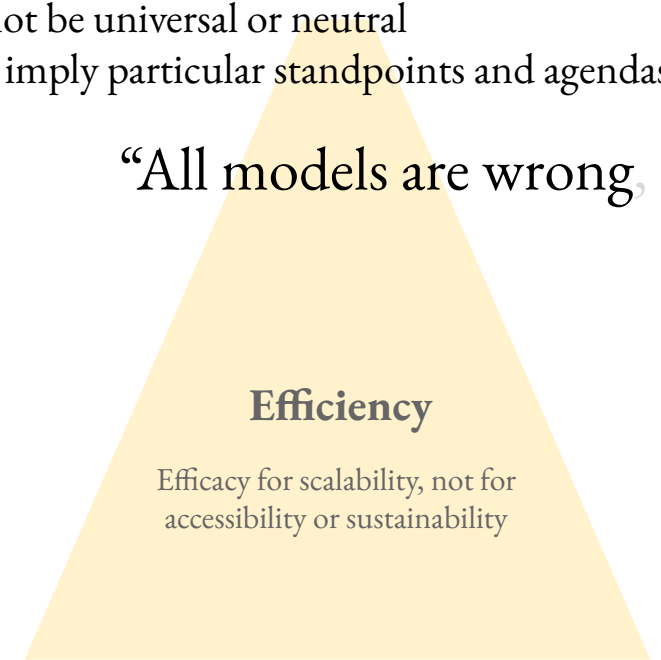



Birhane, et al. 2022. The Values Encoded in Machine Learning Research.

## Case study wrap-up

A seemingly “neutral” and benevolent value statement
may not be universal or neutral
It can imply particular standpoints and agendas

“All models are wrong, but some are useful.”




Birhane, et al. 2022. The Values Encoded in Machine Learning Research.

## Case study wrap-up

A seemingly “neutral” and benevolent value statement
may not be universal or neutral
It can imply particular standpoints and agendas

“All models are wrong, but some are useful.”

If models are all wrong but wrong in different ways,
then it matters to ask: to whom are they useful? For whose interests and benefits?

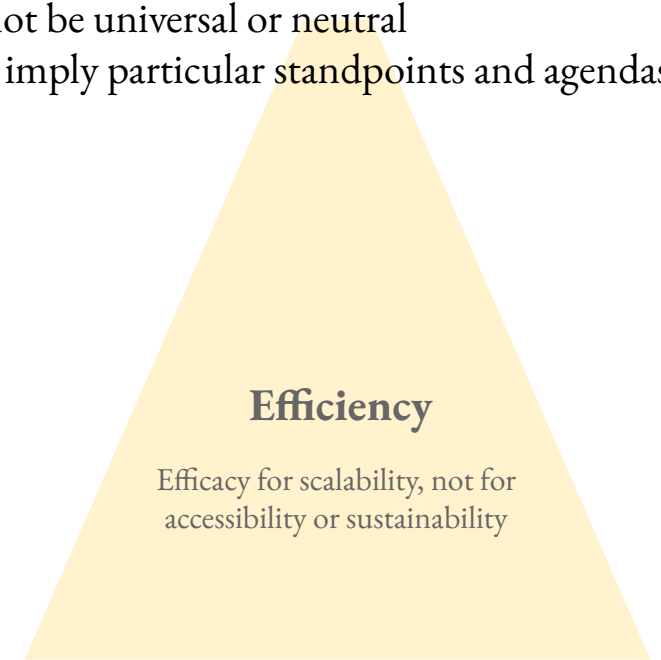
**Efficiency**

Efficacy for scalability, not for accessibility or sustainability

Birhane, et al. 2022. The Values Encoded in Machine Learning Research.

## Lost opportunities in technical agendas

A seemingly “neutral” and benevolent value statement may not be universal or neutral
It can imply particular standpoints and agendas

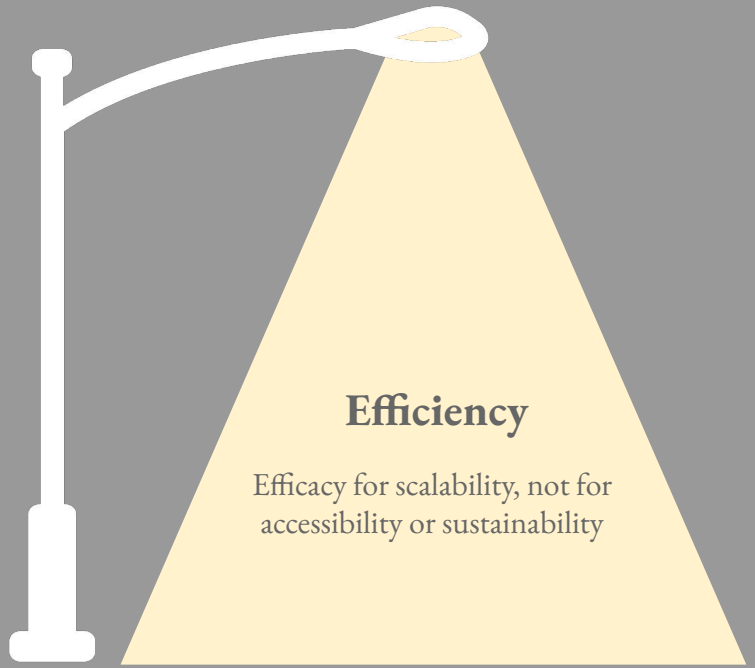



Efficiency that “can facilitate work by low-resource communities”

Efficiency that “can lessen resource extraction – e.g. less hardware or data harvesting or lower carbon emissions”

Birhane, et al. 2022. The Values Encoded in Machine Learning Research.

## Streetlight effect




## Lost opportunities in technical agendas

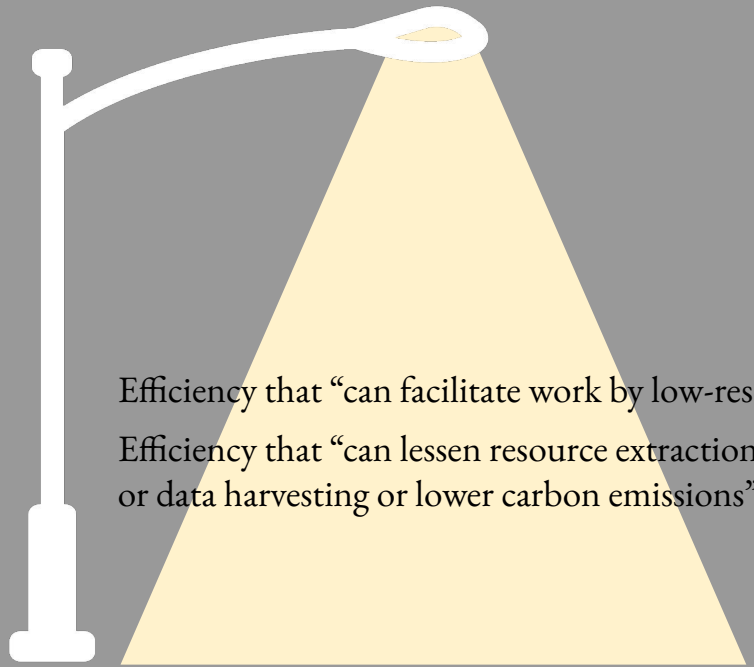
Efficiency that "can facilitate work by low-resource communities"

Efficiency that "can lessen resource extraction – e.g. less hardware or data harvesting or lower carbon emissions"

Birhane, et al. 2022. The Values Encoded in Machine Learning Research.

## Streetlight effect

## Lost opportunities in technical agendas



**Efficiency**

Efficacy for scalability, not for accessibility or sustainability

Efficiency that "can facilitate work by low-resource communities"

Efficiency that "can lessen resource extraction – e.g. less hardware or data harvesting or lower carbon emissions"

Birhane, et al. 2022. The Values Encoded in Machine Learning Research.

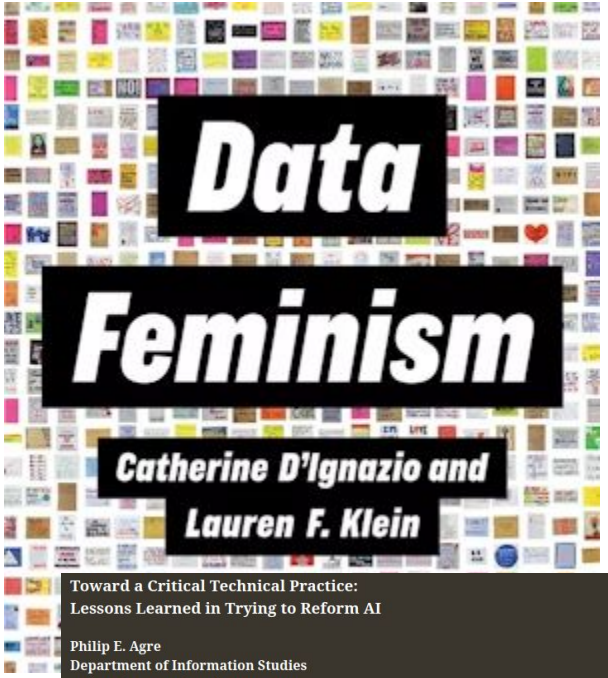
# Critical Research and Critical Technical Practice

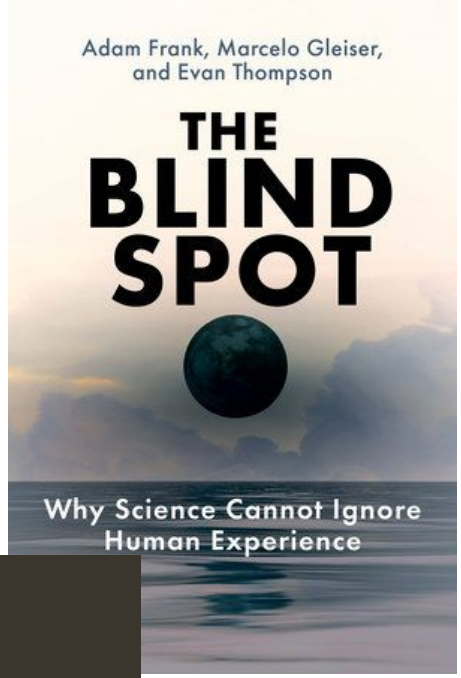
1. Can the answers to the “why” question really be justified?
2. Critical research and critical technical practice is to ask deep, critical questions to understand the layers of layers of frames that set the background assumptions of taken-for-granted beliefs.

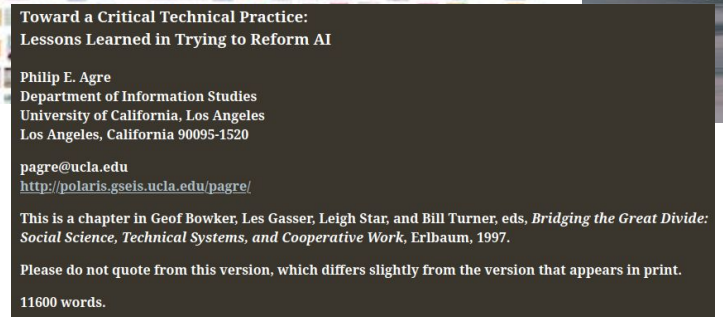
| | |
|---|---|
| **1. Problem formulation** | **1a. Appropriateness of problem formulation**<br>Justify why the problem formulation of the technique is appropriate, and why this problem should be addressed by rather than other alternative techniques or non-technical approaches. |
| | **1b. Who can benefit from the development of the technique?**<br>Identify who will directly benefit from proposing the technique, and justify how problem motivation and formulation of the technique stem from stakeholders', users', and/or the public's interests. |
| | **1c. Who can potentially be put at risk due to the development of the technique?**<br>Describe which population may be potentially harmed, placed at risk, excluded, or overlooked, due to the development of the technique. |

| | |
|---|---|
| **2. Model design** | **2a. Design assumptions and domain knowledge**<br>Describe the domain knowledge and justify any assumptions incorporated in the model design. |
| **3. Model evaluation** | **3a. Specifying the scope and level of evidence for the evaluation method**<br>Report the scope and level of evidence for the evaluation method. |
| **4. Limitation analysis** | **4a. Limitation assessment**<br>Conduct thorough assessments to understand the scope, weaknesses, harms, and failure modes of the model. |
| | **4b. Disclosing technique-specific limitations**<br>Disclose the model-specific limitations in its problem formulation, design assumptions, and evaluation methods. |
| | **4c. Disclosing the intrinsic limits of the modeling approach and the evaluation methods**<br>Disclose the intrinsic limits that cannot be overcome with technological progress to avoid ignoring harms caused by their inherent defects. |

# Readings





Toward a Critical Technical Practice:
Lessons Learned in Trying to Reform AI

Philip E. Agre
Department of Information Studies
University of California, Los Angeles
Los Angeles, California 90095-1520

pagre@ucla.edu
http://polaris.gseis.ucla.edu/pagre/

This is a chapter in Geof Bowker, Les Gasser, Leigh Star, and Bill Turner, eds, *Bridging the Great Divide: Social Science, Technical Systems, and Cooperative Work*, Erlbaum, 1997.

Please do not quote from this version, which differs slightly from the version that appears in print.

11600 words.

# Critical Computing Communities

## Critical Computing:

https://criticalcomputing.cc

## Critical AI:

https://read.dukeupress.edu/critical-ai

## Critical data studies:

https://en.wikipedia.org/wiki/Critical_data_studies

## Our website and slack:

https://critical-technical-practice.github.io/

# Thank you!

Critical Research and Critical Technical Practice

A tutorial

---

Any critical comments, feedbacks, and suggestions are welcome!

Critical Technical Practice Working Group @ SFU critical-technical-practice.github.io

Contact: Weina Jin (weinaj@sfu.ca)