

# Rigor in AI

Alexandra Olteanu

Microsoft Research Montreal

## ***Rigor in AI: Doing Rigorous AI Work Requires a Broader, Responsible AI-Informed Conception of Rigor***

---

Alexandra Olteanu<sup>1\*</sup> Su Lin Blodgett<sup>1</sup> Agathe Balayn<sup>1</sup> Angelina Wang<sup>2</sup>

Fernando Diaz<sup>3</sup> Flavio du Pin Calmon<sup>4</sup> Margaret Mitchell<sup>5</sup>

Michael Ekstrand<sup>6</sup> Reuben Binns<sup>7</sup> Solon Barocas<sup>1</sup>

<sup>1</sup>Microsoft Research <sup>2</sup>Cornell Tech <sup>3</sup>Carnegie Mellon University <sup>4</sup>Harvard University

<sup>5</sup>Hugging Face <sup>6</sup>Drexel University <sup>7</sup>University of Oxford

\*Corresponding author: alexandra.olteanu@microsoft.com

### **Abstract**

In AI research and practice, *rigor* remains largely understood in terms of *methodological rigor*—such as whether mathematical, statistical, or computational methods are correctly applied. We argue that this narrow conception of rigor has contributed to the concerns raised by the responsible AI community, including overblown claims about the capabilities of AI systems. Our position is that a broader conception of what rigorous AI research and practice should entail is needed. We believe such a conception—in addition to a more expansive understanding of (1) *methodological rigor*—should include aspects related to (2) what background knowledge informs what to work on (*epistemic rigor*); (3) how disciplinary, community, or personal norms, standards, or beliefs influence the work (*normative rigor*); (4) how clearly articulated the theoretical constructs under use are (*conceptual rigor*); (5) what is reported and how (*reporting rigor*); and (6) how well-supported the inferences from existing evidence are (*interpretative rigor*). In doing so, we also provide useful language

Impoverished notions of rigor do not only have one-off undesirable outcomes but can have a deeply formative impact on the quality of both AI research and practice.

# Rigor in AI research and practice

Often understood in terms of **methodological rigor**, often conceptualized as a mix of e.g.,

- whether mathematical, statistical or computational methods are correctly applied
- whether new methods/models/systems are tested on large or complex benchmarks and compared with a sufficient number of other competing methods, models or systems
- whether the methods or analyses can scale
- whether the phenomena under analysis were mathematically formalized and quantified (in contrast to more qualitative work)

# What Responsible AI has to do with rigor?

Responsible AI is about minimizing and preventing harms

At the same time, by asking for

- better documentation practices
- better evaluation and measurement practices
- better development practices
- better understanding of possible adverse impacts
- better problem formulation and for conceptual clarity
- better engagement with stakeholders
- better consent practices
- ...

responsible AI is also asking for more rigor in AI research and practice

# Facets of Rigor in AI research and practice

**Epistemic rigor** – what background knowledge informs what to work on?

**Normative rigor** - how do norms, standards, and values influence the work?

**Conceptual rigor** - are theoretical constructs under investigation clearly articulated?

**Methodological rigor** - are methodological choices appropriate and sound?

**Reporting rigor** - what research findings are reported and how they are reported?

**Interpretative rigor** - how well supported are the inferences and any claims being made by existing evidence?

# Facets of Rigor in AI research and practice

Facet of rigor	What the facet is concerned with	What the facet asks for
<b>Epistemic rigor</b> (§2.1)	What <i>background knowledge</i> informs which problems are addressed and how?	Is the <i>background knowledge</i> clearly and explicitly communicated? Is the <i>background knowledge</i> appropriate, well-justified, and appropriately applied?
<b>Normative rigor</b> (§2.2)	Which disciplinary, community, organizational, or personal <i>norms, standards, values, or beliefs</i> influence the work and how?	Are these <i>norms, standards, values, or beliefs</i> clearly and explicitly communicated? Are these <i>norms, standards, values, or beliefs</i> appropriate & appropriately followed?
<b>Conceptual rigor</b> (§2.3)	Which <i>theoretical constructs</i> are under investigation?	Are the <i>theoretical constructs</i> clearly and explicitly articulated? Are the <i>theoretical constructs</i> appropriate and well-justified?
<b>Methodological rigor</b> (§2.4)	Which <i>methods</i> are being used?	Are these <i>methods</i> and their use clearly and explicitly described? Are these <i>methods</i> appropriate, well-justified and appropriately applied?
<b>Reporting rigor</b> (§2.5)	What are the <i>research findings</i> ?	Are the <i>research findings</i> clearly communicated? Is the presentation of <i>research findings</i> appropriate and well-justified?
<b>Interpretative rigor</b> (§2.6)	What <i>inferences</i> are being drawn from the research findings?	Are these <i>inferences</i> clearly and explicitly communicated? Are these <i>inferences</i> appropriate and well-justified?

All facets of rigor are inherently about the choices we make about an object of concern (e.g., *epistemic rigor* is concerned with *background knowledge*, while *conceptual rigor* is concerned with *theoretical constructs*)



# Facets of Rigor in AI research and practice

Facet of rigor	What the facet is concerned with	What the facet asks for
Epistemic rigor (§2.1)	What <i>background knowledge</i> informs which problems are addressed and how?	Is the <i>background knowledge</i> clearly and explicitly communicated? Is the <i>background knowledge</i> appropriate, well-justified, and appropriately applied?
Normative rigor (§2.2)	Which disciplinary, community, organizational, or personal <i>norms, standards, values, or beliefs</i> influence the work and how?	Are these <i>norms, standards, values, or beliefs</i> clearly and explicitly communicated? Are these <i>norms, standards, values, or beliefs</i> appropriate & appropriately followed?
Conceptual rigor (§2.3)	Which <i>theoretical constructs</i> are under investigation?	Are the <i>theoretical constructs</i> clearly and explicitly articulated? Are the <i>theoretical constructs</i> appropriate and well-justified?
Methodological rigor (§2.4)	Which <i>methods</i> are being used?	Are these <i>methods</i> and their use clearly and explicitly described? Are these <i>methods</i> appropriate, well-justified and appropriately applied?
Reporting rigor (§2.5)	What are the <i>research findings</i> ?	Are the <i>research findings</i> clearly communicated? Is the presentation of <i>research findings</i> appropriate and well-justified?
Interpretative rigor (§2.6)	What <i>inferences</i> are being drawn from the research findings?	Are these <i>inferences</i> clearly and explicitly communicated? Are these <i>inferences</i> appropriate and well-justified?

All facets of rigor are inherently about the choices we make about an object of concern (e.g., *epistemic rigor* is concerned with *background knowledge*, while *conceptual rigor* is concerned with *theoretical constructs*)

# Facets of Rigor in AI research and practice

Facet of rigor	What the facet is concerned with	What the facet asks for
<b>Epistemic rigor</b> (§2.1)	What <i>background knowledge</i> informs which problems are addressed and how?	Is the <i>background knowledge</i> clearly and explicitly communicated? Is the <i>background knowledge</i> appropriate, well-justified, and appropriately applied?
<b>Normative rigor</b> (§2.2)	Which disciplinary, community, organizational, or personal <i>norms, standards, values, or beliefs</i> influence the work and how?	Are these <i>norms, standards, values, or beliefs</i> clearly and explicitly communicated? Are these <i>norms, standards, values, or beliefs</i> appropriate & appropriately followed?
<b>Conceptual rigor</b> (§2.3)	Which <i>theoretical constructs</i> are under investigation?	Are the <i>theoretical constructs</i> clearly and explicitly articulated? Are the <i>theoretical constructs</i> appropriate and well-justified?
<b>Methodological rigor</b> (§2.4)	Which <i>methods</i> are being used?	Are these <i>methods</i> and their use clearly and explicitly described? Are these <i>methods</i> appropriate, well-justified and appropriately applied?
<b>Reporting rigor</b> (§2.5)	What are the <i>research findings</i> ?	Are the <i>research findings</i> clearly communicated? Is the presentation of <i>research findings</i> appropriate and well-justified?
<b>Interpretative rigor</b> (§2.6)	What <i>inferences</i> are being drawn from the research findings?	Are these <i>inferences</i> clearly and explicitly communicated? Are these <i>inferences</i> appropriate and well-justified?

All facets of rigor are inherently about the choices we make about an object of concern (e.g., *epistemic rigor* is concerned with *background knowledge*, while *conceptual rigor* is concerned with *theoretical constructs*)



# Why are these distinctions useful

These facets are not intended to prescribe certain choices, but to ask us to think carefully about and make those choices explicit

They provide distinct lenses for reflecting on the quality and scientific integrity of AI research and practice

Untangle debates about these different choices

# What does a focus on methodological rigor overlooks?

The work and the claims we make are shaped by a variety of choices that both precede and succeed any methodological considerations



# Conceptual rigor

Which *theoretical constructs* are under investigation? Are these *theoretical constructs* clearly and explicitly articulated? Are these *theoretical constructs* appropriate and well-justified?

**Example:** Assume a researcher wishes to evaluate whether a LLM “hallucinates.”

In AI research, the construct of “hallucination” has been used to refer to several distinct types of system behaviors, including generating text that is

- Nonsensical
- Factually incorrect
- Not in the input data
- Not in the training data
- ...

# Rigor in AI

Alexandra Olteanu

Microsoft Research Montreal

## ***Rigor in AI: Doing Rigorous AI Work Requires a Broader, Responsible AI-Informed Conception of Rigor***

---

Alexandra Olteanu<sup>1\*</sup> Su Lin Blodgett<sup>1</sup> Agathe Balayn<sup>1</sup> Angelina Wang<sup>2</sup>

Fernando Diaz<sup>3</sup> Flavio du Pin Calmon<sup>4</sup> Margaret Mitchell<sup>5</sup>

Michael Ekstrand<sup>6</sup> Reuben Binns<sup>7</sup> Solon Barocas<sup>1</sup>

<sup>1</sup>Microsoft Research <sup>2</sup>Cornell Tech <sup>3</sup>Carnegie Mellon University <sup>4</sup>Harvard University

<sup>5</sup>Hugging Face <sup>6</sup>Drexel University <sup>7</sup>University of Oxford

\*Corresponding author: alexandra.olteanu@microsoft.com

### **Abstract**

In AI research and practice, *rigor* remains largely understood in terms of *methodological rigor*—such as whether mathematical, statistical, or computational methods are correctly applied. We argue that this narrow conception of rigor has contributed to the concerns raised by the responsible AI community, including overblown claims about the capabilities of AI systems. Our position is that a broader conception of what rigorous AI research and practice should entail is needed. We believe such a conception—in addition to a more expansive understanding of (1) *methodological rigor*—should include aspects related to (2) what background knowledge informs what to work on (*epistemic rigor*); (3) how disciplinary, community, or personal norms, standards, or beliefs influence the work (*normative rigor*); (4) how clearly articulated the theoretical constructs under use are (*conceptual rigor*); (5) what is reported and how (*reporting rigor*); and (6) how well-supported the inferences from existing evidence are (*interpretative rigor*). In doing so, we also provide useful language