

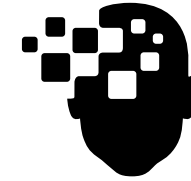


Position:

Stop Treating ‘AGI’ as the North-star Goal of AI Research

Borhane Bili-Hamelin, Christopher Graziul, Leif Hancox-Li, Hananel Hazan, El-Mahdi El-Mhamdi, Avijit Ghosh, Katherine Heller, Jacob Metcalf, Fabricio Murai, Eryk Salvaggio, Andrew Smart, Todd Snider, Mariame Tighanimine, Talia Ringer, Margaret Mitchell, Shiri Dori-Hacohen

ARGUMENT



ICML
International Conference
On Machine Learning

PROBLEM

- **Definitions of artificial general intelligence (AGI) are highly contested!**
- Contested ideas aren't a problem, unless people ignore these tensions
- Current discourse about AGI ignores tensions, exacerbating hype

POSITION

- **Pursuit of AGI catalyzes known issues in AI via (at least) six traps!**
- This obscures worthwhile scientific, engineering, and societal goals
- We offer three recommendations for avoiding these traps

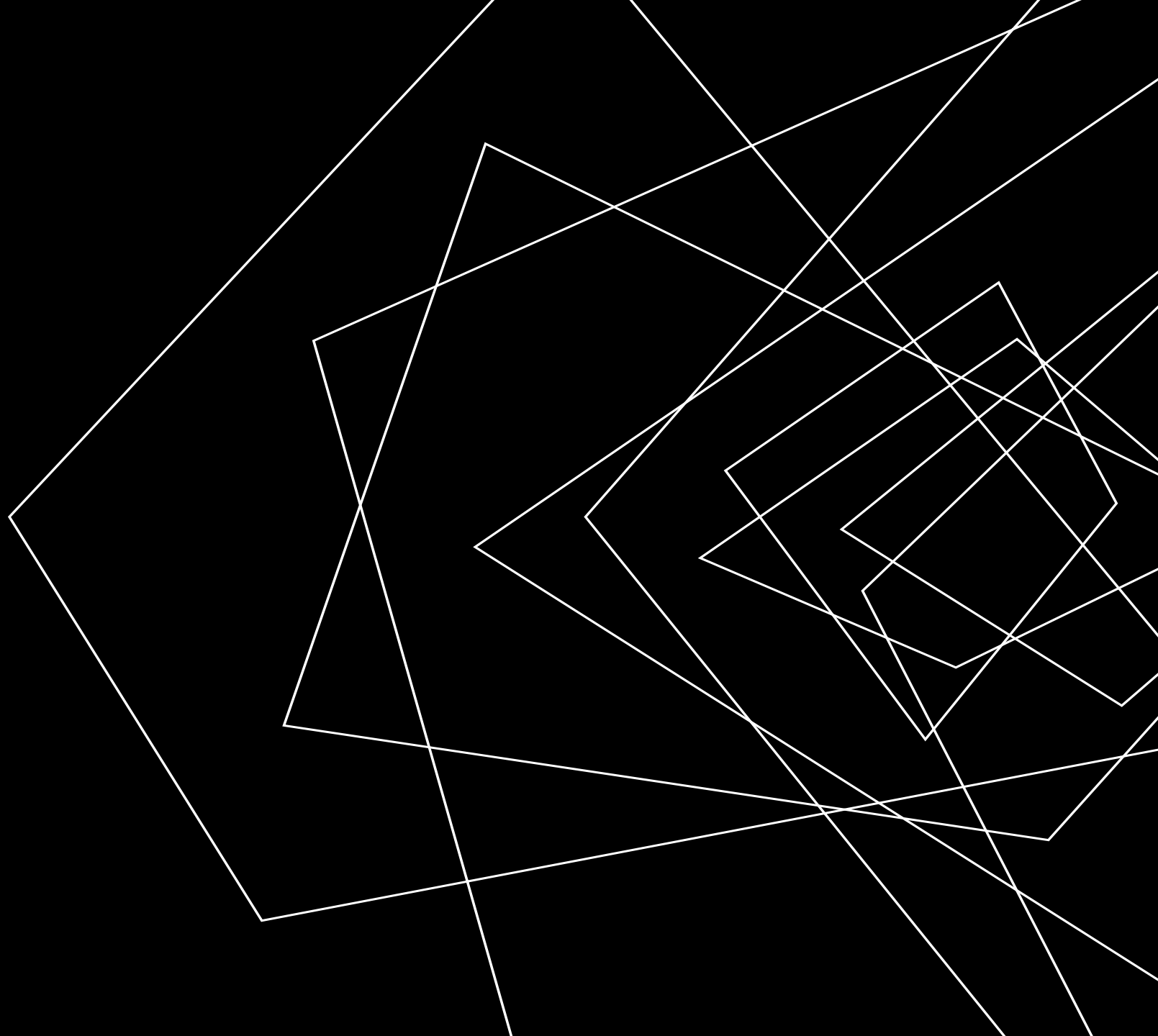
OVERVIEW

Argument

Traps

Recommendations

Takeaway



TRAPS

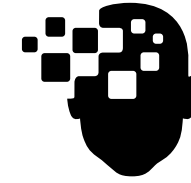
1. ILLUSION OF CONSENSUS

- **Using shared term(s) in a way that gives a false impression of consensus about goals, despite goals being contested**
- Obstructs explicit engagement with what AI research goals actually are and should be

2. SUPERCHARGING BAD SCIENCE

- **Worsening current problems with bad science in AI due to poorly defined concepts and experimental procedures**
- Ex: When you can't define your target, you can't design proper studies or evaluation metrics

TRAPS



ICML
International Conference
On Machine Learning

3. PRESUMING VALUE-NEUTRALITY

- Framing goals as purely technical or scientific, when they are in fact laden with political, social, or ethical values
- Sidesteps normative questions about which research directions merit pursuit

4. GOAL LOTTERY

- Adopting goals which are not adequately justified by scientific, engineering, or social merit, but instead on the basis of incentives, circumstances, or luck
- Ex: Allowing hype cycles and marketing claims about what seems AGI-like to drive research agendas

TRAPS

5. GENERALITY DEBT

- **Relying on the generality or flexibility of tools to postpone crucial engineering, scientific, or societal decisions**
- Tradeoff between vague pursuit of generality and user-relevant capabilities

6. NORMALIZED EXCLUSION

- **Excluding communities and experts from shaping goals**
- Undermines innovation and collective legitimacy by excluding the very communities and experts needed to define worthwhile research goals

RECOMMENDATIONS

R1. SPECIFICITY

Be *highly specific* when describing scientific, engineering, and societal goals

R2. PLURALISM

Allow for *many* worthwhile scientific, engineering, and societal goals

R3. INCLUSION

Make space at the table, it supports innovation and collective legitimacy

TAKEAWAY

The future of AI is too important to leave to a vague three-letter acronym

THE REAL GOAL

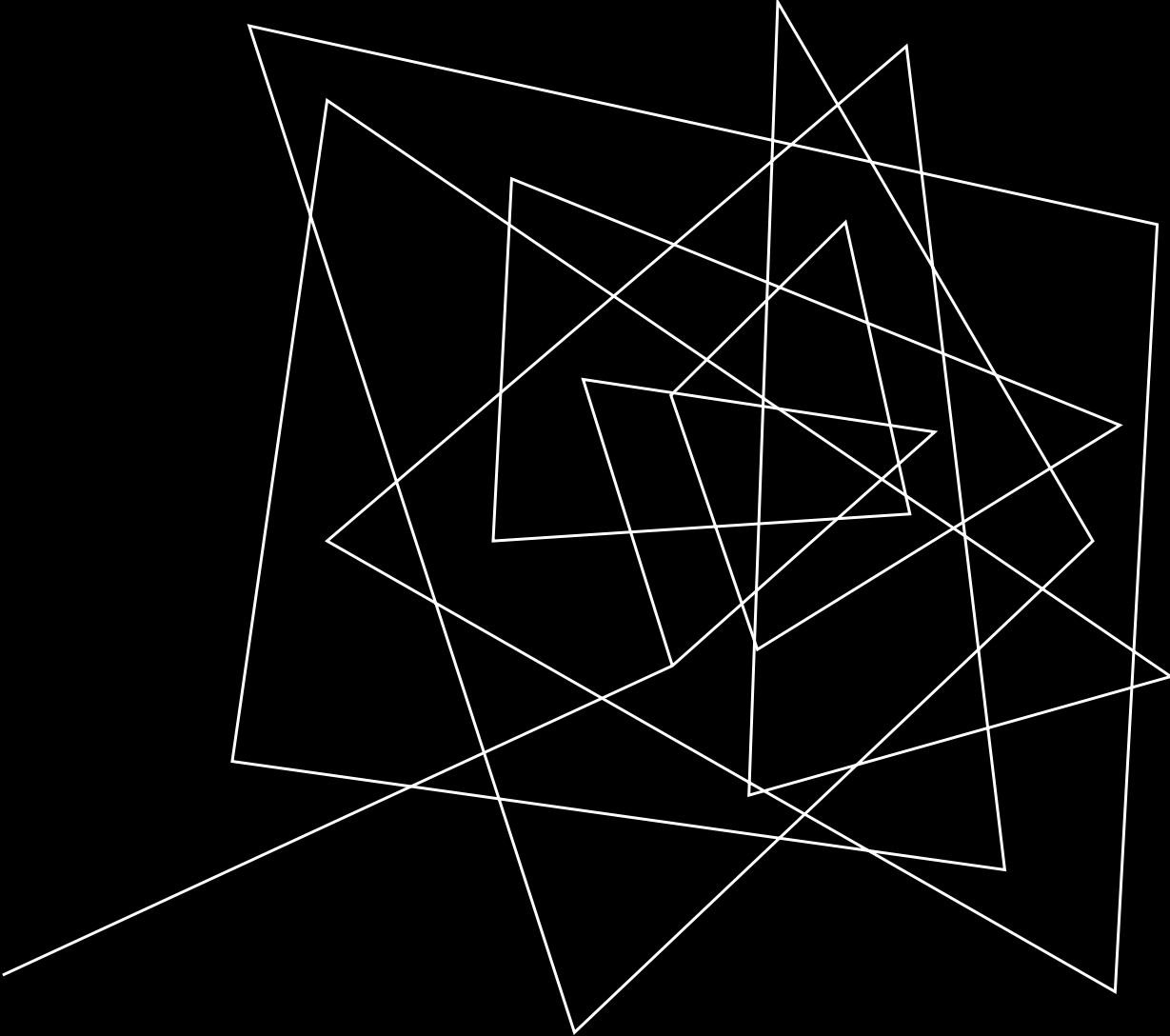
If we need a unifying vision, let it be supporting and benefiting human beings

STOP

Chasing vague goals that privilege hype over good science

START

Embracing research practices that help distinguish hype from reality



THANK YOU!

ACKNOWLEDGMENTS

Borhane Blili-Hamelin was funded in part through a Magic Grant from the Brown Institute for Media Innovation. Christopher Graziul was supported by the National Institute of Minority Health and Health Disparities of the National Institutes of Health under award number R01MD015064. This material is based upon work supported in part by the NSF Program on Fairness in AI in Collaboration with Amazon under Award IIS-2147305.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, National Institutes of Health, the Brown Institute for Media Innovation, or Amazon.