

# Values-Centered Adoption of AI for the USGCRP

Katie Shilton

# Mapping Human/Organizational Values to AI

Collaboration with colleagues in computer science, philosophy, and information

One outcome: heuristics and exercises for making AI adoption decisions that center human values.

- 1) Mapping principles to AI capabilities and challenges
- 2) Mapping AI breakdowns and avenues for repair



VALUES-CENTERED  
ARTIFICIAL INTELLIGENCE

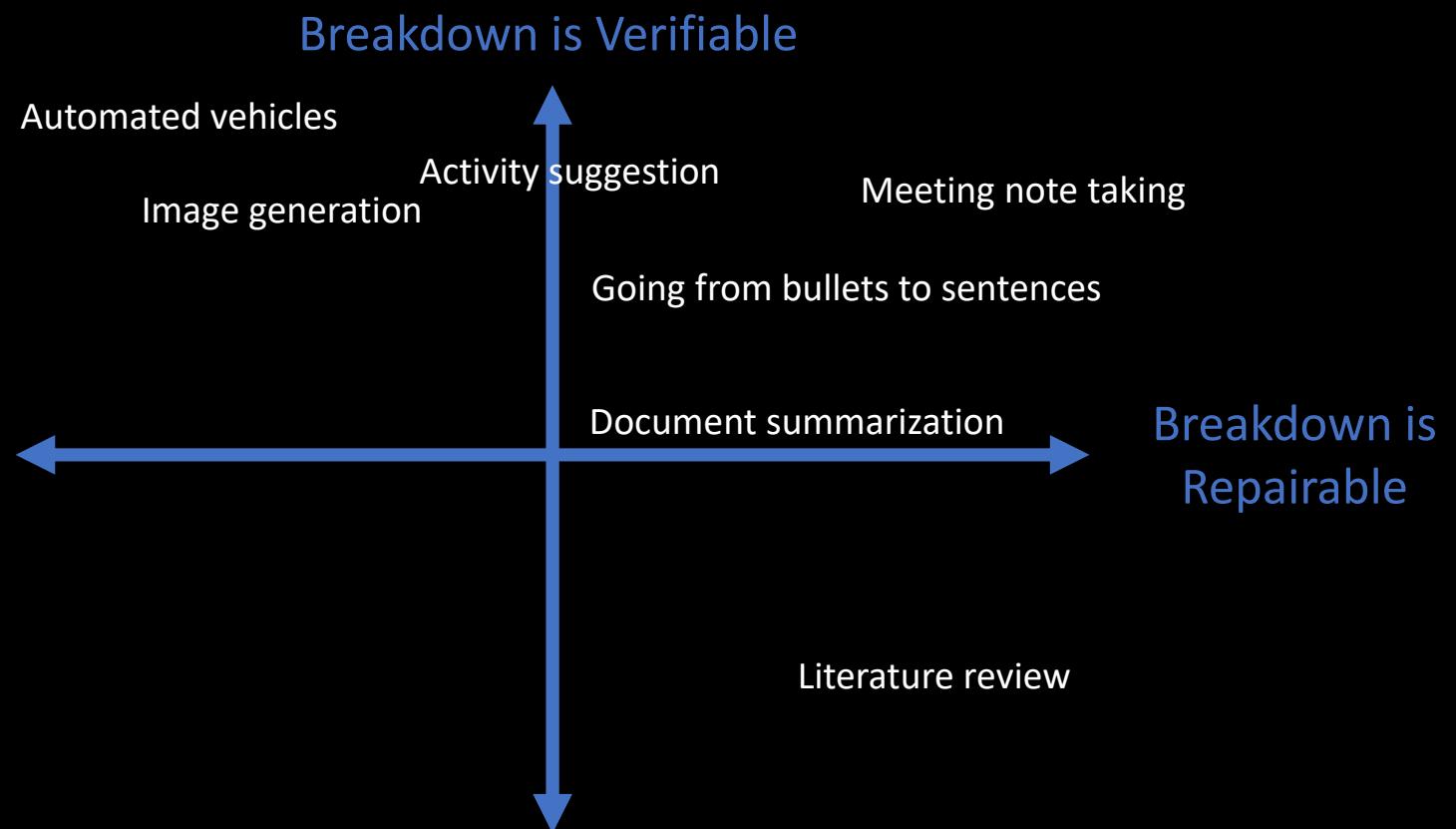
# Mapping USGCRP Principles

| Principle  | GenAI supports by... | GenAI challenges by.... |
|--|----------------------|-------------------------|
| Applicability and Utility: source is important, relevant, and useful for its intended audience   |                      |                         |
| Transparency and Traceability: source material, methods, and evaluation are documented and clear.  |                      |                         |
| Objectivity: purpose, methods, presentation, substance, and interpretation of conclusions are clear, accurate, reliable, and unbiased.                         |                      |                         |
| Integrity and Security: source material will remain protected and intact over time; both info and owners of info are respected.                                |                      |                         |
| Reproducibility: Procedures surrounding source materials are documented such that they can be reproduced, with checks for robustness on non-reproducible data. |                      |                         |

# Mapping USGCRP Principles

| Principle  | GenAI supports by...  | GenAI challenges by....  |
|--|---|--|
| Applicability and Utility: source is important, relevant, and useful for its intended audience   | <ul style="list-style-type: none"><li>*Potentially* expanding access to peer review, grey lit, expert &amp; local knowledge.</li><li>Translation of indigenous languages.</li></ul> | <ul style="list-style-type: none"><li>Obscuring sources</li></ul>                                |
| Transparency and Traceability: source material, methods, and evaluation are documented and clear.  | <ul style="list-style-type: none"><li>Potential improvements for documentation of processes</li></ul>   | <ul style="list-style-type: none"><li>Lacking provenance and traceability</li></ul>              |
| Objectivity: purpose, methods, presentation, substance, and interpretation of conclusions are clear, accurate, reliable, and unbiased.                         |   | <ul style="list-style-type: none"><li>May introduce language or representation biases.</li></ul> |
| Integrity and Security: source material will remain protected and intact over time; both info and owners of info are respected.                                |   | <ul style="list-style-type: none"><li>Remixing, disrespecting ownership</li></ul>                |
| Reproducibility: Procedures surrounding source materials are documented such that they can be reproduced, with checks for robustness on non-reproducible data. |   | <ul style="list-style-type: none"><li>Producing variable results</li></ul>                       |

# Mapping AI Breakdown and Repair



# Values-Centered Adoption Tools

- Algorithmic Impact Assessment tools: OECD: <https://oecd.ai/en/catalogue/tools/algorithmic-impact-assessment-tool>, Ada Lovelace (healthcare): <https://www.adalovelaceinstitute.org/resource/aia-user-guide/> EqualAI: <https://www.equalai.org/aia/>
- Vakkuri, V., Kemell, K.-K., Jantunen, M., Halme, E., & Abrahamsson, P. (2021). ECCOLA — A method for implementing ethically aligned AI systems. *Journal of Systems and Software*, 182, 111067. <https://doi.org/10.1016/j.jss.2021.111067>

# Questions for discussion

- What AI tools are we considering, for what purposes?
  - Which are classifiers? Which are generative AI?
- What are their points of breakdown, and how – and who- will repair?
- How do those tools map to agency and process values?
  - And which stakeholders should be consulted?

Thank you!

[kshilton@umd.edu](mailto:kshilton@umd.edu)

TRAILS

<https://www.trails.umd.edu/>

Ethics & Values in Design (EViD) Lab

<http://evidlab.umd.edu/>



*This material is based on work supported by National Science Foundation under awards 1704369 and 2229885.*